

Second Edition

ALYSSA NEY

# metaphysics

an introduction



# METAPHYSICS

*Metaphysics: An Introduction, Second Edition* combines comprehensive coverage of the core elements of metaphysics with contemporary and lively debates within the subject. It provides a rigorous and yet accessible overview of a rich array of topics, connecting the abstract nature of metaphysics with the real world. Topics covered include:

- basic logic for metaphysics
- an introduction to ontology
- abstract objects
- material objects
- critiques of metaphysics
- natural and social kinds
- the metaphysics of race and gender
- grounding and fundamentality
- free will
- time
- modality
- persistence
- causation.

This outstanding book not only equips the reader with a thorough knowledge of the fundamentals of metaphysics, but provides a valuable guide to contemporary metaphysics and metaphysicians.

New for this second edition are updated sections on metaphysical indeterminacy; new sections on manipulation arguments and free will and neuroscience; and new chapters on the important topics of social ontology, fundamentality, and grounding.

Additional features such as exercises, annotated further reading, a glossary, and a companion website <https://routledgetextbooks.com/textbooks/9780815350491> have also been updated and will help students find their way around this subject and assist teachers in the classroom.

**Alyssa Ney** is a Professor of Philosophy at the University of California - Davis, USA. She works primarily in metaphysics, philosophy of mind, and philosophy of physics. She is the author of *The World in the Wave Function: A Metaphysics for Quantum Physics* (2021), and co-editor with David Z Albert of *The Wave Function: Essays on the Metaphysics of Quantum Mechanics* (2013).

### **Praise for the first edition**

"An up-to-date, well-written text that is both challenging and accessible. I think that the greatest strengths of the book are its science-friendliness and the interweaving of under-represented issues, such as social construction, race, and numbers, with traditionally-favoured topics."

–Matthew Slater, *Bucknell University, USA*

"Exemplary clarity and concision ... The author has presented some difficult material in a light, brisk and appealing style."

–Barry Lee, *University of York, UK*

"An excellent introduction ... some very complicated and important work is made accessible to students, without either assuming background knowledge or over-simplifying."

–Carrie Jenkins, *University of British Columbia, Canada*

"This book will serve as an excellent introduction to contemporary metaphysics. The issues and arguments are well chosen and explained with a carefulness and rigor ideal for beginning students. The teaching-oriented materials, which include a helpful overview of elementary logic, are useful additions to Ney's expert discussion."

–Sam Cowling, *Denison University, USA*

"Over the last several decades, metaphysics has been a particularly active and productive area of philosophy. Alyssa Ney's *Metaphysics: An Introduction* offers a superb introduction to this exciting field, covering the issues, claims, and arguments on fundamental topics, such as existence and persistence, material object, causation, modality, and the nature of metaphysics. While the presentation is admirable in its clarity and accessibility, Ney does not shy away from sophisticated problems and theories, many of them from recent developments in the field, and she succeeds in infusing them with immediacy and relevance. The reader has a sense of being a fellow companion on Ney's journey of exploration into some fascinating metaphysical territories. This is the best introduction to contemporary metaphysics that I know."

–Jaegwon Kim, *Brown University, USA*

"This is a terrific text. In a remarkably short space Alyssa Ney manages to be simultaneously comprehensive, authoritative and deep. She gives a cutting-edge account of all the standard topics, and for good measure adds an illuminating discussion of the metaphysics of race. This will be not only be a boon to students, but also a valuable resource for more experienced philosophers."

–David Papineau, *King's College London, UK*

"One of the best introductions to metaphysics available. It covers a wide range of contemporary topics in metaphysics, giving a clear, accessible yet substantive account of the key questions and issues and providing an up to date account of the current debate. It's the text I'd choose for my own course on the subject."

–L.A. Paul, *University of North Carolina, Chapel Hill, USA*

# METAPHYSICS

## An Introduction

*Second Edition*

Alyssa Ney



Cover image: Jay's photo / Getty Images

Second edition published 2023

by Routledge

4 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge

605 Third Avenue, New York, NY 10158

*Routledge is an imprint of the Taylor & Francis Group, an informa business*

© 2023 Alyssa Ney

The right of Alyssa Ney to be identified as authors of this work has been asserted by them in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

First edition published by Routledge, 2014

*British Library Cataloguing-in-Publication Data*

A catalogue record for this book is available from the British Library

ISBN: 978-0-8153-5048-4 (hbk)

ISBN: 978-0-8153-5049-1 (pbk)

ISBN: 978-1-351-14120-8 (ebk)

DOI: 10.4324/9781351141208

Typeset in Interstate

by Apex CoVantage, LLC

Access the Instructor and Student Resources: <https://routledgetextbooks.com/textbooks/9780815350491>

This book is dedicated to the memory of my father, Garrett Ney.

# CONTENTS

<i>Preface</i>	xi
<i>Notes on the Second Edition</i>	xiii
<i>Acknowledgments</i>	xiv
<i>Visual Tour of Metaphysics: An Introduction</i>	xv
<b>Preparatory Background: Logic for Metaphysics</b>	<b>1</b>
0.1 Arguments	1
0.2 Validity	4
0.3 Soundness	8
0.4 Criticizing Arguments	9
0.5 The Principle of Charity and Enthymemes	11
0.6 Propositional Logic	14
0.7 First-Order Predicate Logic	18
Suggestions for Further Reading	28
<b>1 An Introduction to Ontology</b>	<b>29</b>
1.1 Ontology: A Central Subfield of Metaphysics	29
1.2 The Puzzle of Nonexistent Objects	30
1.3 Finding One's Ontological Commitments: Quine's Method	35
1.4 The Method of Paraphrase	41
1.5 Ockham's Razor	46
1.6 Where Should Metaphysical Inquiry Begin?	48
Suggestions for Further Reading	51
<b>2 Abstract Entities</b>	<b>54</b>
2.1 More Than a Material World?	54
2.2 The Abstract/Concrete Distinction	56
2.3 Universals and the One Over Many Argument	59
2.4 Applying Quine's Method	65
2.5 Nominalism and Other Options	71
2.6 Mathematical Objects	76
Suggestions for Further Reading	81
<b>3 Material Objects</b>	<b>84</b>
3.1 What is a 'Material' Object?	84
3.2 The Paradoxes of Material Constitution	86

3.3 The Problem of the Many	96
3.4 The Special Composition Question	99
3.5 Moderate Answers to the Special Composition Question	100
3.6 Mereological Nihilism	105
3.7 Mereological Universalism	107
3.8 Vagueness	110
3.9 Back to the Paradoxes	112
Suggestions for Further Reading	113
<b>4 Critiques of Metaphysics</b>	116
4.1 A Concern about Methodology	116
4.2 Carnap's Two Critiques of Metaphysics	117
4.3 Responses to Carnap's Arguments	125
4.4 Present Day Worries about Metaphysical Method	130
4.5 The Relationship Between Metaphysics and Science: A Proposal	132
Suggestions for Further Reading	134
<b>5 Natural and Social Kinds</b>	136
5.1 A Traditional Account of Natural Kinds	136
5.2 The Homeostatic Property Cluster Account	141
5.3 Social Kinds and Social Reality	143
5.4 Social Construction	145
5.5 Constructing Human Kinds	147
5.6 Causal vs. Constitutive Construction	153
5.7 What Are We Doing When We Do Social Ontology?	154
Suggestions for Further Reading	156
<b>6 The Metaphysics of Race and Gender</b>	158
6.1 Three Views about the Metaphysics of Race and Gender	158
6.2 Biological Realism about Race	159
6.3 Social Constructionism or Eliminativism?	164
6.4 Working Out a Social Constructionism about Race	166
6.5 Gender and Sex	169
6.6 Biological Realism about Gender	171
6.7 Essentialism or Eliminativism?	174
6.8 Social Constructionism about Gender	177
6.9 Conferralism about Gender	180
6.10 A World Without Gender?	181
Suggestions for Further Reading	182
<b>7 Grounding and Fundamentality</b>	184
7.1 Fundamentality Questions	184
7.2 The Mind-Body Problem	185
7.3 Supervenience	190
7.4 The Limits of Supervenience	194
7.5 Grounding	197
7.6 Fundamentality	200
7.7 Fundamentality as Ontological Independence	202

7.8 The Fundamental as a Basis for the Nonfundamental	210
7.9 Is There a Fundamental Level?	213
Suggestions for Further Reading	216
<b>8 Time</b>	219
8.1 Time's Passage	219
8.2 Background on Special Relativity	224
8.3 The Argument from Special Relativity Against Presentism	229
8.4 The A-Theory and the B-Theory	232
8.5 The Truthmaker Objection	237
8.6 Time Travel	248
Suggestions for Further Reading	253
<b>9 Persistence</b>	256
9.1 The Puzzle of Change	256
9.2 Some Views about Persistence	259
9.3 A Solution to the Paradoxes of Material Constitution	261
9.4 The Problem of Temporary Intrinsics	264
9.5 Exdurantism	269
9.6 Defending Three-Dimensionalism	272
Suggestions for Further Reading	275
<b>10 Modality</b>	277
10.1 Possibility and Necessity: Modes of Truth	277
10.2 Varieties of Possibility and Necessity	278
10.3 The Possible Worlds Analysis of Modality	280
10.4 Ersatz Modal Realism	289
10.5 Rejecting the Possible Worlds Analysis	294
10.6 Essentialism and Anti-Essentialism	298
10.7 Essentialism Today	300
Suggestions for Further Reading	303
<b>11 Causation</b>	305
11.1 Causation in the History of Philosophy	305
11.2 Hume's Empiricism	308
11.3 Three Reductive Theories of Causation	311
11.4 An Objection to Reductive Theories of Causation	320
11.5 Process Theories of Causation	322
11.6 Two Projects in the Philosophy of Causation	325
Suggestions for Further Reading	326
<b>12 Free Will</b>	329
12.1 What is Free Will?	329
12.2 The Problem of Free Will and Determinism	331
12.3 Determinism	334
12.4 Compatibilism	337
12.5 Frankfurt Cases	342
12.6 Manipulation Arguments	343
12.7 Libertarianism	347

x *Contents*

12.8 Skepticism about Free Will	351
Suggestions for Further Reading	354

<i>Glossary of Terms</i>	357
<i>Works Cited</i>	371
<i>Index</i>	382



# PREFACE

The distinctive goal of the metaphysician is to understand the structure of reality: what kinds of entities exist and what are their most fundamental and general features and relations. Unlike the natural and social sciences that seek to describe some special class of entities and what they are like – the physical things or the living things, particular civilizations or cultures – metaphysicians ask the most general questions about how things are, what our universe is like.

We will have more to say in the chapters that come about what are the main issues in metaphysics today and what exactly is the relationship between metaphysics and those other ways we have of studying what the world is like, science, and theology. In this preface, our aim is to orient the reader with a basic overview of the presentation, and supply some suggestions for further resources that will complement the use of this textbook.

This book presents an introduction to contemporary analytical metaphysics aiming to be accessible to students encountering the topic for the first time, and yet challenging and interesting to more advanced students who may have already seen some of these topics in a first philosophy course. To say this book presents an introduction to contemporary *analytical* metaphysics is to signal that the emphasis of this book will be in stating views and arguments clearly and with logical precision. As a result, in many places this book will make use of the tools of modern symbolic logic. Ideally a student using this book will already have had a course introducing the basics of first-order predicate logic. For those who have not already had such a course, a preparatory chapter is provided which should bring one up to speed. This chapter may also be useful as a review to students who have already seen this material, or may be skimmed to find the notation that is used throughout the remainder of the text.

This textbook contains several features that have been included to help the introductory student who may be encountering many of these concepts for the first time. This includes a glossary at the end of the book, as well as a list of suggested readings accompanying each chapter. The aim of the glossary, it should be noted, is not to provide philosophical analyses

of terms or views. These are in many cases up for debate in contemporary metaphysics. The aim of the glossary is merely to give a gloss of the relevant term or view that will be helpful to orient a reader. Terms in the text that have glossary entries are marked in **boldface type**.

In addition to the suggested readings at the end of each chapter, there are also several excellent general resources that are available. Students planning to write papers on any of the topics in this book would do well to consult the following websites and handbooks:

- The *Stanford Encyclopedia of Philosophy* and the *Internet Encyclopedia of Philosophy* are two free, online encyclopedias. All articles are written by professional philosophers.
- [www.philpapers.org](http://www.philpapers.org) is a free website cataloguing published and unpublished articles and books in philosophy. In addition to including a searchable database of works in philosophy, this website also provides useful bibliographies on a variety of topics.
- The journal *Philosophy Compass* publishes survey articles on many topics in contemporary philosophy aimed at an advanced undergraduate/beginning graduate student audience.

In addition to these online resources, the following two books in metaphysics provide useful introductions to many of the topics we discuss here and beyond:

- The *Oxford Handbook of Metaphysics*, edited by Michael Loux and Dean Zimmerman.
- Blackwell's *Contemporary Debates in Metaphysics*, edited by John Hawthorne, Theodore Sider, and Dean Zimmerman.

The website accompanying this textbook provides links to many of the articles discussed in these chapters as well as selections from the further reading lists.

Although much of this introduction concerns contemporary metaphysics, the topics and debates that are most discussed today, and the various methodologies that are most common now, it is often useful to recognize the contribution of philosophers and scientists of the past. This book adopts the convention of noting the years of birth and death for all deceased philosophers discussed in the main body of the text. If no dates are provided, one should assume that this philosopher is still living and writing.

# NOTES ON THE SECOND EDITION

Since the first edition of *Metaphysics* was published in 2014, the field has evolved in several ways. Many changes have been made to the second edition to reflect this evolution. These notes are intended to inform instructors who have used the first edition of the main changes they will find in the second.

First, in the past decade, social ontology has taken a much more central place in the field. For this reason, a new chapter on natural and social kinds has been added to the text (Chapter 5). This chapter also includes (Section 5.7) a section on methodology, which introduces Strawson's influential distinction between descriptive and prescriptive metaphysics, as well as Haslanger's ameliorative approach to engaging metaphysical questions. In addition, the chapter on the metaphysics of race from the first edition has been rewritten and expanded so that it now comprises an introduction to both the metaphysics of race and gender (Chapter 6).

A new chapter on Grounding and Fundamentality has also been added to the text (Chapter 7). This replaces the earlier section at the end of Chapter 1 on the topic. This new chapter introduces the concepts of supervenience and grounding, and raises questions both about what it could mean to say something is fundamental and whether anything is in fact absolutely fundamental. The inclusion of this additional chapter is intended to reflect how so many metaphysical questions are today formulated – not merely as questions about what exists, but about what is (absolutely or relatively) fundamental.

Instructors will also notice some other significant changes in the text, also reflecting developments in the field. First, the chapter on material objects now includes a discussion of ontic vagueness (Section 3.8), as new work has made this concept far less obscure and contentious than it was at the time the previous edition was written. Finally, the free will chapter has been updated to include a discussion of both Manipulation Arguments (Section 12.6) and Libet experiments (Section 12.8). The chapter still includes discussion of the Consequence Argument; however, these additional sections were added to reflect what now occupies more of the contemporary debate.

# ACKNOWLEDGMENTS

I wish to thank several people who have helped to make this book actual. First, I would like to thank Tony Bruce of Routledge, who first raised the idea to me of writing this text. I thank Tony for his encouragement and seeing the project through. I would also like to thank Adam Johnson of Routledge for his support and patience as I completed the second edition. I am grateful to Jordan Bell, who updated all of the web content for the second edition. Thanks to Allan Hazlett for his contribution to the chapter on race in the first edition and the use of the Borges example which is retained in the second edition. I wish to thank Karen Bennett, Sam Cowling, Daniel Nolan, Alison Peterman, David Smith, and Peter van Inwagen for comments on the first edition, and Mark Balaguer, Jordan Bell, Ned Markosian, Michael McKenna, Alejandro Naranjo Sandoval, Jonathan Schaffer, Ted Sider, Hannah Tierney, and three anonymous referees for reading portions of the second edition and providing helpful comments. I am really fortunate to work in a field with so many brilliant and generous colleagues. They are not responsible for any flaws that remain in the text. Thanks to my metaphysics students and teaching assistants at the University of Rochester and UC Davis, and to the teachers who first got me passionate about metaphysics and released me from the spell of logical positivism, especially Jose Benardete from whom I took my first metaphysics course, John Hawthorne, Jaegwon Kim, Ted Sider, and Bob Van Gulick. I was not the first and will certainly not be the last student to realize that so many of the questions they had thought were questions for their physics classes were metaphysical questions and that the philosophy department was where I belonged!

# VISUAL TOUR OF METAPHYSICS: AN INTRODUCTION

## Learning Points

At the beginning of each chapter, a number of Learning Points are set out so that the student understands clearly what is to be covered in the forthcoming chapter.

### Learning Points

- Introduces the concept of an argument and procedures for assessing arguments as valid or invalid, sound or unsound
- Provides students with tools for recognizing incomplete arguments (enthymemes) and applying the principle of charity
- Presents basic notation and valid inference forms in propositional and first-order predicate logic

## Emboldened Glossary Terms

A Glossary at the back of the book helps with new terms and their definitions. Where these terms are used for the first time in the book they can be found in **bold** and in the margin.

### Glossary of Terms

**A posteriori method:** an empirical way of knowing a fact or proposition, one that involves observation or sensory experience

**A priori method:** a way of knowing a fact or proposition that does not involve observation or sensory experience

**Abstract:** a classification of entities, examples include properties and mathematical objects

## Exercises

Each chapter includes Exercises that students can undertake inside or outside the class. These give students an opportunity to assess their understanding of the material under consideration.

### EXERCISE 1.1

#### The Argument for Nonexistent Entities

Put the argument for nonexistent entities presented in the second paragraph of this section in numbered premise form. Before seeing Quine's own response to this argument, which premise or premises of this argument do you think one should consider rejecting?

## Annotated Reading

At the end of each chapter, there are Suggestions for Further Reading with annotations explaining their context.

### Suggestions for Further Reading

There are many excellent critical thinking and introductory logic textbooks available that develop the material introduced in this chapter further. Some excellent critical thinking texts are Richard Feldman's *Reason and Argument* and Thomas McKay's *Reasons, Explanations, and Decisions: Guidelines for Critical Thinking*. Some excellent introductory logic texts are Merrie Bergmann, James Moor, and Jack Nelson's *The Logic Book* and Gary Hardegree's *Symbolic Logic: A First Course*.

# Preparatory Background

## Logic for Metaphysics

### Learning Points

- Introduces the concept of an argument and procedures for assessing arguments as valid or invalid, sound or unsound
- Provides students with tools for recognizing incomplete arguments (enthymemes) and applying the principle of charity
- Presents basic notation and valid inference forms in propositional and first-order predicate logic

### 0.1 Arguments

In metaphysics, as in most other branches of philosophy and the sciences, we are interested in finding the truth about certain topics. For this reason, it would be nice to have an accessible and reliable method to arrive at the truth. We aren't going to find what is true by random guessing. And in philosophy, we don't think that the best method to find the truth is to simply trust what one has always believed, those views one was raised with (though common sense should be respected to some extent). Nor do we think there is a group of elders who have the truth so that the correct method of discovery is just to seek them out and find what they have said.<sup>1</sup> Instead what we do is seek out arguments for various positions: series of statements that rationally support a particular position and can allow us to see for ourselves why a position is correct. It is because philosophers want a trustworthy method for arriving at the truth that much of our time is spent seeking out good arguments.

The word 'argument' has a specific meaning in philosophy that is different from its ordinary usage. When we say 'argument,' we don't mean two people yelling at each other. Also, we should emphasize since this is a common confusion, that when we say 'argument,' we don't simply mean one person's position or view. Rather an **argument** is typically a series

**Argument:** a series of statements in which someone is presenting reasons in defense of some claim



## 2 Preparatory Background

**Premise:** a statement offered as part of an argument as a reason for accepting a certain claim

**Conclusion:** the part of an argument that is being argued for, for which reasons are being offered

**Theism:** the thesis that God exists

**Atheism:** the thesis that God does not exist

**Numbered premise form:** a way of stating arguments so that each premise, as well as the conclusion, are given a number and presented each on their own line

of statements presenting reasons in defense of some claim. Most arguments have two components. First, they have **premises**. These are the statements that are being presented as the reasons for accepting a certain claim. Second, they have a **conclusion**. This is the claim that is being argued for, the statement for which reasons are being given. Here are examples of some metaphysical arguments you might have seen in your first philosophy class:

*The Argument from Design* (for **theism**: the thesis that God exists)

The complexity and organization of the universe shows that it must have been designed. But there cannot be something which is designed without there being a designer. So, the universe must have a designer. Therefore, God exists.

*The Problem of Evil* (for **atheism**: the thesis that God does not exist)

If there were a God, He would not allow evil to exist in this world. But there is evil in this world. Therefore, God does not exist.

Each set of statements constitutes an argument because there is a claim being defended, a conclusion, and reasons being offered in defense of that claim, the premises. To better reveal the structure of an argument, throughout this book, we will often display arguments in the following form, numbering the premises and conclusion. We will call this **numbered premise form**. Here is how we might present the Argument from Design in numbered premise form:

*The Argument from Design*

1. The complexity and organization of the universe shows that it must have been designed.
2. But there cannot be something designed without there being a designer.
3. So, the universe must have a designer.

Therefore,

4. God exists.

And similarly for the Problem of Evil:

*The Problem of Evil*

1. If there were a God, He would not allow evil to exist in this world.
2. But there is evil in this world.

Therefore,

3. God does not exist.

When we present arguments this way, it allows us to refer easily back to the premises, and if we are interested in criticizing the argument, to single out which ones are questionable or in need of more defense.

In the two examples we have just now considered, it is quite easy to figure out which are the premises and which is the conclusion. Sometimes in a text it is more difficult to figure out which is which, or to figure out in which order one should state the premises. The following exercises will help you work through some more challenging cases. One tool that will help you get these arguments into numbered premise form is to look for words that typically signal a premise or a conclusion.

- *Words and phrases that tend to indicate premises:* since, for, because, due to the fact that, ...
- *Words and phrases that tend to indicate conclusions:* hence, thus, so, therefore, it must be the case that, ...

You will then want to organize the premises in such a way that they naturally lead to the conclusion.

### EXERCISE 0.1

#### Recognizing Premises and Conclusions

The following paragraphs include the kinds of arguments that were presented in the United States in 2009 for and against nationalized healthcare. Decide which sentences are the premises and which is the conclusion in each case, and state the argument in numbered premise form. Note that the conclusion may not be presented last in the argument.

- A. Americans should reject nationalized healthcare. This is because a system with nationalized healthcare is one in which someone's parents or baby will have to stand in front of the government's death panel for bureaucrats to decide whether they are worthy of healthcare. Any system like that is evil.
- B. If we don't nationalize healthcare, there may be those, especially the young and healthy, who will take the risk and go without coverage. And if we don't nationalize healthcare, there will be companies that refuse to give their workers coverage. When people go without coverage, the rest of the country pays for them. So, if the young and healthy or employees go without coverage, then the rest of the country will have to pay more in taxes. No one should have to pay more in taxes. Therefore, we should nationalize healthcare.

## 0.2 Validity

What we would like in philosophy is to find good arguments that present us with compelling reasons to believe their conclusions are true. This comes down to two issues. First, we want to find arguments that have premises that are independently reasonable to believe. Second, we want to find arguments whose premises logically imply their conclusions. This means if we have reason to believe the premises are true, we will have reason to believe the conclusion is true as well.

What we are going to do in the first part of this chapter is provide you with tools that will allow you to articulate clearly in what way a certain argument is a good argument or a bad argument other than just by simply stating, "That argument is good," "I like that argument," or "That's a bad argument," "I don't like that argument." When debating important topics at a high level, we want to be more articulate than that, and these next few sections will give you the vocabulary to be so.<sup>2</sup>

One important feature we look for in a good argument is that it be a valid argument. 'Validity' is a technical term referring to a logical feature of an argument. By definition, an argument is (**deductively**<sup>3</sup>) **valid** just in case there is no way for its premises to all be true while its conclusion is false. In other words, in a valid argument, if the premises are all true, then the conclusion must also be true. In valid arguments, we say the conclusion "follows deductively" from the premises. An argument is (**deductively**) **invalid** if it is possible for all of the premises of the argument to be true while the conclusion is false. In an invalid argument, the truth of the premises does not guarantee the truth of the conclusion.

When we speak about validity, I will emphasize again, this is a logical feature of an argument. It is all about whether the conclusion can be said to logically follow from the premises. It is not about whether the premises of an argument are as a matter of fact true. It is only about whether *if* the premises were true, the conclusion would *also* have to be true. The question about the truth of the premises is of course important and it is something we will discuss in the next section. It is just not what we care about when we are interested in validity. Let's run through a few examples of arguments to illustrate this definition of validity.

### *Argument 1*

1. If the universe were to end tomorrow, we would never know if there exist alien life forms.
2. The universe will not end tomorrow.

Therefore,

3. We will get to know if there exist alien life forms.

What should we say about this argument? Is this argument valid or invalid? To assess this, all we need to do is ask ourselves the following question:

**Deductively valid:** an argument is deductively valid when there is no way for the premises to all be true while the conclusion is false. The premises of the argument logically imply its conclusion

**Deductively invalid:** an argument is deductively invalid when it is possible for the premises of the argument to all be true while its conclusion is false

Is it possible for there to be a situation in which the premises of this argument are all true and yet its conclusion is false? This is what you should ask yourself every time you are asked to assess the validity of an argument. This is because it follows from the definitions of validity and invalidity that if it turns out there is a possible scenario, one that doesn't involve a contradiction, in which the premises are all true and yet the conclusion is false, this is an *invalid* argument. We are not talking about a likely situation, just one we can understand that doesn't commit us to something of the form  $p$  and not- $p$  (this is all we mean by a **contradiction**). If the premises could all be true while the conclusion is false, then the conclusion doesn't follow logically from the premises. And so, by definition, the argument is invalid. Otherwise, it is valid.

**Contradiction:** any sentence or statement of the form  $p$  and not- $p$

So what we should do to assess Argument 1's validity is try to see if we can understand how the following situation could obtain without a contradiction:

<b>TRUE</b>	If the universe were to end tomorrow, we would never know if there exist alien life forms.
<b>TRUE</b>	The universe will not end tomorrow.
<b>FALSE</b>	We will get to know if there exist alien life forms.

Can we tell a story in which this is the case? Could it be that the first two statements are both true and yet the third is false? Yes, this is easy to see. We start by supposing that (1) is true. We haven't yet discovered alien life forms, and so if the universe were to end tomorrow, we would never know if there are any. Then we imagine it is also true that the universe does not end tomorrow. This doesn't rule out the conclusion being false: that even though the universe doesn't end tomorrow, we still never get to learn whether there are alien life forms. Perhaps we never learn this because the universe ends next week rather than tomorrow. Since there is a coherent situation in which both premises are true and yet the conclusion is false, the argument is invalid.

In general, when you describe a situation in which the premises of an argument are true and the conclusion is false, what you are doing is providing a **counterexample** to the argument.

Let's try this with another case:

#### *Argument 2*

1. All events have a cause.
2. The Big Bang is an event.

Therefore,

3. The Big Bang has a cause.

What should we say about the validity of this argument? Remember: validity is a logical property of an argument. It is not about whether the premises of an argument are in fact true, but whether they as a matter of logic

**Counterexample:** an example that shows an argument is invalid, by providing a way in which the premises of the argument could be true, while the conclusion is false; or an example that shows a universal statement is false, by providing an instance in which it is false

6 Preparatory Background

deductively entail their conclusion. So to assess this argument's validity, we should set aside any skepticism we might have about the *actual* truth of the premises themselves. We just want to know in the *possible* (though perhaps not actual) scenario where the premises are true, could the conclusion be false.

So, is this a valid argument? To settle this again all we need to do is see whether there is a possible situation in which the premises are all true and the conclusion is false. And again, by 'possible,' we mean *logically possible*. We are asking: is this a situation we can imagine, one that involves no contradiction, in which the premises are all true and the conclusion false?

Here, it turns out: no. There is no possible situation in which the premises of this argument are both true and yet this conclusion is false.

<b>TRUE</b>	All events have a cause.
<b>TRUE</b>	The Big Bang is an event.
<b>FALSE</b>	The Big Bang has a cause.

Once we fix the premises and make them true, the conclusion has to be true too. If *all* events have a cause and the Big Bang is an event, then the Big Bang must have a cause too. To assume the conclusion is false is to assume the Big Bang does not have cause. So, a situation in which the premises are true and the conclusion is false is one in which the Big Bang both is and is not an event – a contradiction. Since there is no possible situation in which the premises are true and the conclusion is false, the above argument is *valid*. This doesn't mean the above argument is good in every way. There may be some other negative things to say about it. For example, one might be skeptical of the actual truth of one or more of its premises. But at least in terms of its logic, this is a good argument; it is valid.

Table 0.1 illustrates one key point that you should draw from this section: the question of an argument's validity is independent of the actual truth or falsity of its premises and conclusion. There can be invalid arguments with all true premises and a true conclusion. And there can be valid arguments with all false premises and a false conclusion. All that matters for validity is the logical connection *between* the premises and the conclusion.

This table shows all possible combinations of premises and conclusion. As you can see, there is only one combination that can never occur. You will never find a valid argument in which the premises are all actually true and the conclusion is actually false. This follows from the definition of validity: a valid argument is one in which there is no possible way for the premises to all be true while the conclusion is false.

Table 0.1

<p><b>Premises: All true</b>  <b>Conclusion: True</b>  <b>Valid Argument</b></p> <p>1. If Paris is in France, then it is in Europe.  2. Paris is in France.  Therefore,  3. Paris is in Europe</p> <p><b>Invalid Argument</b></p> <p>1. If Paris is in France, then it is in Europe.  2. Paris is in Europe.  Therefore,  3. Paris is in France.</p>	<p><b>Premises: All true</b>  <b>Conclusion: False</b>  <b>Valid Argument</b></p> <p><i>It is not possible to have a valid argument with all true premises and a false conclusion.</i></p> <p><b>Invalid Argument</b></p> <p>1. If Paris is in Spain, then it is in Europe.  2. Paris is in Europe.  Therefore,  3. Paris is in Spain.</p>
<p><b>Premises: At least one false</b>  <b>Conclusion: True</b>  <b>Valid Argument</b></p> <p>1. If Paris is in China, then it is in Europe.  2. Paris is in China.  Therefore,  3. Paris is in Europe.</p> <p><b>Invalid Argument</b></p> <p>1. If Paris is in France, then it is in Asia.  2. Paris is in Asia.  Therefore,  3. Paris is in France.</p>	<p><b>Premises: At least one false</b>  <b>Conclusion: False</b>  <b>Valid Argument</b></p> <p>1. If Paris is in Spain, then it is in Asia.  2. Paris is in Spain.  Therefore,  3. Paris is in Asia.</p> <p><b>Invalid Argument</b></p> <p>1. If Paris is in Spain, then it is in Asia.  2. Paris is in Asia.  Therefore,  3. Paris is in Spain.</p>

## EXERCISE 0.2

### Testing Arguments for Validity

Are the following arguments valid or invalid?

- A. All lawyers like basketball. Barack Obama is a lawyer. Therefore, Barack Obama likes basketball.
- B. Some snakes eat mice. Mice are mammals. Therefore, some snakes eat some mammals.
- C. If the Pope is a bachelor, then the Pope lives in an apartment. The Dalai Lama is a bachelor. So, the Dalai Lama lives in an apartment.
- D. All birds can fly. Penguins are birds. But penguins can't fly. Therefore, some birds can't fly.



### 0.3 Soundness

**Sound:** an argument is sound just in case it has all true premises and is deductively valid

If there is just one thing philosophers are looking for when they seek out good arguments, most would probably say what they are looking for is soundness. An argument is **sound** just in case it has two features. First, it must be a valid argument, in the sense just defined. Second, all of its premises must actually be true. When an argument is sound, it presents good reason to believe its conclusion. This is because by knowing it is sound, we know (i) that *if* its premises are true, its conclusion must be true as well, and (ii) that its premises are, as a matter of fact, true.

In Section 0.2 on validity, we considered two arguments. We can now evaluate whether these are sound arguments. The first we considered, about the universe ending tomorrow and the aliens, fails to be sound because it is invalid. The second, about the Big Bang, one might think also fails to be sound, but not because it is invalid. Rather one might think the second argument is unsound because it has at least one false premise. Here is an example of a sound argument:

#### *Argument 3*

1. Greece is a member of the European Union.
2. All members of the European Union lie north of the Equator.

Therefore,

3. Greece lies north of the Equator.

This is a sound argument because it satisfies both conditions: (i) it is valid, and (ii) it has all true premises. We can check to see that it is valid by using the method in the previous section. We see if we can coherently imagine a situation in which all of its premises are true while its conclusion is false:

<b>TRUE</b>	Greece is a member of the European Union.
<b>TRUE</b>	All members of the European Union lie north of the Equator.
<b>FALSE</b>	Greece lies north of the Equator.

We can't do that though. To imagine that would involve imagining a contradiction obtaining, Greece being both north of the Equator and not north of the Equator. So, the argument is valid. Since its premises are both *actually* true, it is also sound.

We are most of the time interested in whether arguments for or against a position are sound. So, in general when you are asked to assess an argument in this course, you should first look for the following:

- Are all of the premises of this argument true? If not, which do you think are false and why?
- Does the conclusion follow from the premises? That is, is the argument valid?

If the answers to these questions are 'yes,' then the argument is sound. The premises are true and the conclusion logically follows from them. So, one has reason to believe the conclusion is true as well.

### EXERCISE 0.3

#### Assessing Arguments for Soundness

Go back to Exercise 0.2 at the end of the Section 0.2 and assess these arguments for soundness.

## 0.4 Criticizing Arguments

Once one understands what we are looking for in metaphysics (sound arguments for the positions that are of interest to us), one can also see how to rationally evaluate these arguments. One always has two options for criticizing an opponent's argument. One can either (i) challenge one of the argument's premises, or, one can (ii) challenge the validity of the argument. Let's briefly discuss each of these in turn.

First, let's again consider the Argument from Design presented in the first section (Section 0.1):

#### *The Argument from Design*

1. The complexity and organization of the universe shows that it must have been designed.
2. But there cannot be something which is designed without there being a designer.
3. So, the universe must have a designer.

Therefore,

4. God exists.

We now have the tools to criticize this argument, if this is something we are interested in doing. One option is to criticize it by challenging one of the premises. The other is to challenge the validity of the inferences made in the argument. There are at least two ways to challenge the truth of premise (1). One could argue that the complexity and organization of the universe has no bearing on whether it was designed. Or one could argue that complexity actually shows the universe lacks a designer (perhaps a designer would prefer a simple universe over one with so much complexity). Alternatively, one might instead challenge premise (2). For example, one could argue that the fact that something is designed doesn't imply the existence of a designer. This would be to get into a debate about what it means to say that something is designed. Either way, if one wants to

## 10 Preparatory Background

deny this argument is sound because premise (1) or (2) is false, one would need to present a compelling reason to think the premise in question is indeed false.

**Minor conclusion:** a statement that is argued for on the way to arguing for an argument's major conclusion

**Major conclusion:** the final conclusion of an argument

Note that since (3) is just supposed to follow from (1) and (2) on the way to the conclusion (4), we call it a **minor conclusion**, as opposed to (4) which we call the **major conclusion** of the argument. In criticizing the argument, one may take issue with the steps leading to (3), but note that since it is not an independent premise of the argument, any objection to (3) will really boil down to either an objection to the independent premises (1) or (2), or an objection to the validity of the inference from (1) and (2) to (3).

There are two inferences in this argument that we may assess for their validity. First, there is the move from (1) and (2) to (3). Then there is the move from (3) to the final, major conclusion (4). Both are places one may try to criticize the argument. Here, what one should do is check both steps for validity. First, is it possible for (1) and (2) to be true, while (3) is false? Here the answer is "No." (1) and (2) do logically imply (3). So, it is not the validity of that step that is mistaken here. On the other hand, it is open for one to challenge the validity of the inference from (3) to (4). One might think there is no contradiction that results from assuming that (3) is true, the universe has a designer, and yet (4) is false, God doesn't exist. Perhaps the universe was designed by someone other than God. This situation would constitute a counterexample to the argument.

Either way, if the argument fails to make all valid inferences, or the argument has premises that are false, the argument will fail to be sound. In this case, it fails to provide a compelling reason to believe its conclusion. Note that one may criticize an argument in this way even if one as a matter of fact believes its conclusion. Not every argument for a true conclusion has to be a good argument.

### EXERCISE 0.4

#### Criticizing Arguments

Consider the following argument for theism.

##### *The Cosmological Argument*

1. Everything that happens in the universe must have a cause.
2. Nothing can be a cause of itself.
3. So, there must exist a first cause.
4. If there is a first cause, then this first cause is God.
5. Therefore, God exists.

Identify which premises are supposed to follow from earlier premises in the argument (as minor or major conclusions). Label the independent premises (i.e. those that are neither major nor minor conclusions). If there are reasons to be skeptical about the truth of any of the independent premises, then state these reasons. Then, evaluate whether the inferences that are made to minor or major conclusions all appear valid. Is the argument sound? Why or why not?

## 0.5 The Principle of Charity and Enthymemes

One thing to keep in the back of your mind as you go about evaluating arguments in metaphysics is that all of us are trying to work together as part of a common enterprise to discover the truth. And so, it is a convention of philosophical debate that one applies what is called the **principle of charity**. What this means is that when it is reasonable, one should try to interpret one's opponent's claims as true and their arguments as valid. For example, if you are reading a text or having a philosophical discussion and someone makes a claim that could easily be interpreted in several ways, some of which are true and some of which are obviously false, the principle of charity recommends that you choose the true way to interpret the author.

Another thing you will find is that some of the time when an author presents an argument in a text, they will present their argument only incompletely. That is, they will present what is called an **enthymeme**. An enthymeme is an argument that is incomplete and invalid as stated, yet although the premises as stated do not logically entail the conclusion, one still has reason to believe the argument the author intended *is* valid. In the case of an enthymeme, an author leaves out some premises because they are simply too obvious to state. Stating them would bore the reader, or insult their intelligence. So, the author leaves them out. The principle of charity compels us in such cases, where it is obvious the author intended these missing premises, and the argument needs them in order to be valid, to fill them in for them.

Here is one example of an enthymeme. Suppose you read in a text an author saying the following:

*Argument against abortion*

Anytime one ends the life of a person, it is murder. Abortion ends the life of a fetus. So, abortion is murder. Therefore, abortion is wrong.

**Principle of charity:** a convention of philosophical debate to, where reasonable, try to interpret one's opponent's claims as true and her arguments as valid

**Enthymeme:** an argument that is incomplete as stated and invalid, although it is easy to supply the missing premises that the argument would need to be valid. In the case of an enthymeme, the author left out the missing premises for fear of boring the reader or insulting their intelligence

## 12 Preparatory Background

One might at first try to state the argument this way in numbered premise form:

### *Argument against abortion*

1. Anytime one ends the life of a person, it is murder.
2. Abortion ends the life of a fetus.
3. So, abortion is murder.

Therefore,

4. Abortion is wrong.

One might then criticize the argument for being invalid. For there are two inferences made in this argument: the first is the move from (1) and (2) to the minor conclusion (3):

### *Inference 1*

1. Anytime one ends the life of a person, it is murder.
2. Abortion ends the life of a fetus.
3. So, abortion is murder. (Minor conclusion)

The second is the inference from (3) to (4):

### *Inference 2*

3. Abortion is murder.

Therefore,

4. Abortion is wrong. (Major conclusion)

One might worry that neither of these inferences, as stated, is deductively valid. In the first case, one could argue that (1) and (2) could be true, but (3) false because, although ending the life of a *person* is murder and abortion ends the life of a fetus, abortion doesn't count as murder because a fetus is not a person. One might argue that the second inference is not valid because it could be the case that abortion is murder and yet abortion is not wrong, because murder is not wrong. (Imagine a world very different from ours where the presence of human life is such a plague that murder is altogether a good thing. Such a world might be very different from ours, but there is no contradiction in this possibility.)

At this point, one may just conclude that this argument against abortion is invalid, and so unsound, and so does not present a compelling reason to think abortion is wrong. However, this response would miss something. Here's why. There is a very simple way to fill in both inferences in this argument using supplementary premises that it is reasonable to think the author assumed. And so a better thing to do would be to grant the author the obvious intermediate steps she intends that would make the argument valid. Then we can make sure we have given the argument the best shot we can.

What are the missing links that will give us a valid argument from the premises to the conclusion? How about this?

*Argument against Abortion*

1. Anytime one ends the life of a person, it is murder.
2. Abortion ends the life of a fetus.
- \*2.5 A fetus is a person (fixes the validity of Inference 1)
- So, abortion is murder.
- \*3.5 Murder is wrong (fixes the validity of Inference 2)

Therefore,

4. Abortion is wrong.

We are allowed, indeed compelled, by the principle of charity to supply the author with premises (2.5) and (3.5) only if it is obvious that these are claims the author intended. And in this case, we will say that the original argument is an enthymeme. It is invalid as stated, but it can easily be made into a valid argument by supplying premises that are obvious the author intended, and may only have left out because they thought they were so obvious.

Note that just because it is often reasonable to reconstruct an author's argument in such a way as to make it valid, this does not mean that we have to accept any argument we ever come across in a text. We still have tools with which to disagree. For although now we can see the above argument as valid, there are several premises whose truth one may take issue with. And this includes the originally unstated premises (2.5 and 3.5) that we added to make the argument valid. All are fair game and open for rational disagreement.

Applying the principle of charity and recognizing enthymemes is a skill that one develops over time as one grapples with more and more philosophical arguments. Exercise 0.5 will help you develop this skill.

### **EXERCISE 0.5**

#### **Supplying Missing Premises**

Some call the ancient Greek philosopher Thales (624 BC-c.546 BC) the first philosopher. Thales is famous for arguing that everything is water. Consider the following texts containing arguments against Thales's thesis. Provide the missing premises that will make the arguments valid.

- A. Water boils at 100 degrees Celcius. Mercury boils at 356 degrees Celcius. Therefore, not everything is water.
- B. There were things that existed in the first seconds immediately after the Big Bang. Water did not come into being until hundreds of thousands of years after the Big Bang. So, not everything is water.



## 0.6 Propositional Logic

We've seen that deciding validity is an important tool in assessing the strength of an argument. But sometimes, when an argument has many premises or its inferences are complicated, it is difficult to assess whether or not an argument is valid using the method we introduced in the section on validity. For this reason, philosophers have developed systems of formal logic, rigorous methods for deciding which forms of argument are or are not valid.<sup>4</sup> Here we will just cover a few basics that will give you tools to tell which argument forms can be trusted to yield valid arguments. These are argument forms that recur throughout the discussions in this book.

First, let's clarify what is meant by an argument form. When we talk about the *form* of an argument, we are talking about the kind of shape or structure an argument has, independent of its specific subject matter. For example, consider the following two arguments:

### *Argument 4*

1. If Elon is human, then he is mortal.
2. Elon is human.

Therefore,

3. He is mortal.

### *Argument 5*

1. If determinism is true, then no one has free will.
2. Determinism is true.

Therefore,

3. No one has free will.

These arguments concern very different topics; their subject matter is distinct. And yet, they have something in common: their logical form. To see this most clearly, logicians will replace the premises and conclusion of an argument with symbols. In the system of logic we are considering now, propositional logic, one chooses upper or lower case letters to represent individual statements or propositions. For example, let's introduce the following symbols to represent the basic propositions that make up the premises and conclusions of Arguments 4 and 5.

H: Elon is human.

M: Elon is mortal.

D: Determinism is true.

N: No one has free will.

In propositional logic, the premises and conclusion of an argument will be represented by either single letters (for the basic or 'atomic' propositions)

Table 0.2

English	Logical/formal notation
<b>And</b>	$\wedge$ , &
Elon is human and Elon is mortal	$H \wedge M$
	$H \& M$
<b>Or</b> (Inclusive or, meaning: either p, q, or both p and q)	$\vee$
Either Elon is human or Elon is mortal.	$H \vee M$
<b>If ... then</b>	$\rightarrow$ , $\supset$
If Elon is human, then he is mortal.	$H \rightarrow M$
	$H \supset M$
<b>Not</b>	$\sim$ , $\neg$
Elon is not human.	$\sim H$
	$\neg H$
<b>If and only if</b> (meaning: both if p then q and if q then p)	$\leftrightarrow$ , $\equiv$
Elon is human if and only if he is mortal.	$H \leftrightarrow M$
	$H \equiv M$

or complex symbols formed out of single letters and some linking symbols, the **logical connectives**. The logical connectives are what are used to build complex propositions out of simpler ones.

The connectives typically recognized in propositional logic are: 'and,' 'if ... then,' 'or,' 'not,' and 'if and only if.' Table 0.2 lists the symbols that are most often used to represent these words in logical notation.

In this book, we will always use ' $\wedge$ ' to symbolize 'and,' ' $\vee$ ' for 'or,' ' $\supset$ ' for 'if ... then,' ' $\neg$ ' for 'not,' and ' $\equiv$ ' for 'if and only if.'

Using this notation, we can now symbolize Arguments 4 and 5:

*Argument 4*

1.  $H \supset M$
2.  $H$

Therefore,

3.  $M$

*Argument 5*

1.  $D \supset N$
2.  $D$

Therefore,

3.  $N$

**Logical connectives:**  
symbols used to build  
complex propositions out  
of simpler ones

**EXERCISE 0.6****Symbolizations in Propositional Logic**

Using the key below, symbolize the following sentences in logical notation.

Key:

I: The universe is infinite.

U: The future is unknown.

O: The future is open.

F: Humans have free will.

- A. Either the universe is infinite or the universe is not infinite.
- B. If humans have free will and the future is open, then the future is unknown.
- C. Humans have free will if and only if the future is open.
- D. It is not the case that either the universe is infinite or the future is open.

Once we symbolize the arguments, their logical structure is more clearly revealed and we can see that they share the same logical form. The form they share is called **modus ponens**.

**Modus ponens:** the  
logical form: If p, then q;  
p; Therefore, q

*Modus Ponens*

1. If p, then q.
2. p

Therefore,

3. q

or, using the notation of propositional logic:

1.  $p \supset q$
2. p

Therefore,

3. q

It doesn't matter which order the premises are written in. Modus ponens is one form of argument that logicians nearly always regard as valid.

**Simplification, modus tollens, and disjunctive syllogism** are three additional argument forms that philosophers generally take to be valid. Note in each case, p and q may stand for any proposition whatsoever, no matter how complex.

*Simplification*

- |                 |                 |
|-----------------|-----------------|
| 1. $p \wedge q$ | 1. $p \wedge q$ |
| Therefore, or   | Therefore,      |
| 2. $p$          | 2. $q$          |

*Modus Tollens*

1.  $p \supset q$   
 2.  $\neg q$

Therefore,

3.  $\neg p$

*Disjunctive Syllogism*

- |                |               |
|----------------|---------------|
| 1. $p \vee q$  | 1. $p \vee q$ |
| 2. $\neg p$ or | 2. $\neg q$   |
| Therefore,     | Therefore,    |
| 3. $q$         | 3. $p$        |

If you find an argument that uses one of these argument forms, classical propositional logic will say that it is valid.

**EXERCISE 0.7****Recognizing Valid Argument Forms in Propositional Logic**

First, symbolize the arguments below using the notation of propositional logic and the key from the previous exercise. Then, decide whether the argument's logical form is (a) modus ponens, (b) simplification, (c) modus tollens, (d) disjunctive syllogism, or (e) none of the above.

- A. Either the future is open or the universe is not infinite. The future is not open. Therefore, the universe is not infinite.
- B. If humans have free will, then the future is open. The future is not open. Therefore, humans don't have free will.
- C. If humans have free will, then the future is open. The future is open. Therefore, humans have free will.
- D. If humans have free will, then the future is open. Humans have free will. Therefore, the future is open.
- E. The future is open and it is unknown. So, the future is unknown.

## 0.7 First-Order Predicate Logic

In Section 0.6, we considered some valid forms of inference in propositional logic. Building upon the foundation of propositional logic, logicians have built more powerful logics, logics that recognize more valid argument forms than propositional logic alone. These logics delve deeper into the structure of our statements, and will be indispensable to representing the views and arguments one encounters in contemporary metaphysics.

For the remainder of this chapter, we will consider first-order predicate logic, initially developed by Gottlob Frege (1848-1925). This will afford us some tools that will be helpful for our discussion of ontology in the next chapters. In later chapters, we will build on this foundation, adding modal and tense operators. But let's start simple. Consider the following argument:

### *Argument 6*

1. Alex respects everyone who loves the Beatles.
2. Betty loves the Beatles.

Therefore,

3. Alex respects Betty.

If we just use the tools of propositional logic, we will not be able to prove this is a valid argument. This is because propositional logic alone implies Argument 6 has the following logical form:

1. A
2. B

Therefore,

3. C

And this is not a valid argument form. We would be forced to symbolize it this way because each proposition (1), (2), and (3) is distinct and none contain the sort of parts ('and,' 'or,' etc.) that would allow us to use the connectives introduced in the previous section.

But the above argument is intuitively valid, and so, to show this using symbolic logic, we need more tools with which to symbolize the argument.<sup>5</sup> First-order predicate logic gives us the relevant tools. The key insight is to recognize that in general we can separate propositions into subjects (or noun phrases) and predicates.

To take a simple case, consider the sentence:

Shaq is tall.

In predicate logic, the symbol for a predicate ('is tall') is always a capital letter. In this case, we will use 'T.' The symbol for the predicate is placed

before the symbol for the subject ('Shaq'). We will use 's' to stand for 'Shaq.' The entire sentence or proposition will then be symbolized in predicate logic in the following way:

Ts.

Similarly, 'Ludwig is a philosopher' could be symbolized as:

Pl.

We might also want to symbolize the sentence:

Shaq admires Ludwig.

This would be:

Asl.

Notice again that the symbol for the predicate (in this case, 'admires') always goes in the front. Here our predicate, 'admires,' is a *two-placed predicate* because it takes two noun phrases as inputs. But of course there exist predicates that take more than two inputs. For example, if you've played the game *Clue*, you've probably stated sentences using predicates like:

'\_\_\_murdered\_\_\_in the\_\_\_using the\_\_\_.'

For example, you might say:

Professor Plum murdered Mr. Body in the kitchen using the candlestick.

This can be symbolized as:

Mpbkc.

One thing that will be especially important in the next chapters is that we are able to represent sentences that make reference to some person(s) or object(s), but without using a name. These are general sentences such as:

Somebody is tall.

Somebody murdered Mr. Body in the kitchen using the candlestick.

or:

Nobody is tall.

There is nothing Professor Plum murdered Mr. Body with in the kitchen.

To represent sentences like this, first-order predicate logic uses **variables** (symbols like  $x$ ,  $y$ ,  $z$ , etc.) and what is called the **existential quantifier**. The existential quantifier is this symbol:

$\exists$

**Variables:** symbols like  $x$ ,  $y$ ,  $z$ , etc. used to stand in for other things in a sentence, called the values of the variable

**Existential quantifier:**  $\exists$ , a symbol of predicate logic. When combined with a variable, it can be used to represent a statement to the effect that something exists that is a certain way

## 20 Preparatory Background

Get used to this backwards and upside-down E. It will be a recurring character throughout this book.

To get familiar with the existential quantifier  $\exists$ , consider the sentence:

Somebody is tall

This is symbolized in first-order predicate logic as:

$$\exists xTx$$

This may be read aloud in any of the following ways:

- There exists an  $x$  such that  $x$  is tall.
- There is at least one  $x$  such that  $x$  is tall.
- Some  $x$  is tall.
- Something is tall.

Or, if we know that our domain of quantification includes only persons (more on domain of quantification momentarily), we may read this as:

- There exists a person  $x$  such that  $x$  is tall.
- There is at least one person  $x$  such that  $x$  is tall.
- Some person  $x$  is tall.
- Somebody is tall.

We can also use the existential quantifier to symbolize the sentence:

Somebody murdered Mr. Body in the kitchen using the candlestick,

as:

$$\exists xMxbkc.$$

We may read this as: "There exists an  $x$  such that  $x$  murdered Mr. Body in the kitchen with the candlestick."

Or, we can represent the sentence:

There is something that Professor Plum murdered Mr. Body with in the kitchen,

as:

$$\exists xMpbkx.$$

Note that the variable ' $x$ ' replaces the name of the object we are quantifying over, the referent of the quantifier phrase 'something' or 'somebody.' In the first case, the 'somebody' refers to the  $x$  that is the murderer, so the variable goes in the first place. In the second sentence, the 'something' refers to the  $x$  that is the murder weapon, so the variable goes in the last place.

We can also represent more complex sentences using the existential quantifier. For example, we can symbolize 'Nothing is tall' as:

$\neg \exists x Tx$  (read as: It is not the case that there exists an  $x$  such that  $x$  is tall.)

To say that there is something that is tall and friendly, we can use the following symbolization:

$\exists x (Tx \wedge Fx)$ ,

where ' $Tx$ ' means  $x$  is tall, and ' $Fx$ ' means  $x$  is friendly.

Or,

There is at least one beautiful eagle on that mountain,

can be symbolized as:

$\exists x ((Bx \wedge Ex) \wedge Mx)$ .

Finally, in some cases, one will find sentences that need more than one variable of quantification. For example, one might want to symbolize in predicate logic the sentence:

Some cats love some dogs.

This sentence has two quantifier phrases. It says both that *there exists some*  $x$  such that  $x$  is a cat, but also that *there exists some*  $y$  such that  $y$  is a dog, and that the cat (the  $x$ ) loves the dog (the  $y$ ). So that we do not confuse which variable is referring to the cat and which the dog, we will use distinct variables  $x$  and  $y$  in the symbolization of this sentence:

$\exists x \exists y ((Cx \wedge Dy) \wedge Lxy)$ ,

which we may read back into English as, "There exists an  $x$  and there exists a  $y$  such that  $x$  is a cat and  $y$  is a dog, and the  $x$  loves the  $y$ ."

Note that in all cases where one uses a variable ( $x$ ,  $y$ ,  $z$ , and so on) as part of a complete sentence, the variable should always be contained within the **scope** of a quantifier. Either it is right next to the quantifier in the sentence, or there should be parentheses reaching from a quantifier and surrounding the occurrence of that variable. Consider the variables in the following two sentences:

$Fx$   
 $\exists x Fx \wedge Gx$

In the first sentence,  $x$  is not contained within the scope of any quantifier, and so this sentence does not express a complete thought. It says ' $x$  is  $F$ ,' where ' $x$ ' does not have any clear meaning. In the second case, the  $x$  in the phrase ' $Fx$ ' is contained within the scope of a quantifier, but the ' $x$ ' in the

#### Scope (of a quantifier):

the part of a sentence containing the variables the quantifier is binding.

In symbolic logic, the scope of a quantifier is either the part of the sentence immediately after the quantifier phrase (in a simple sentence like ' $\exists x Fx$ ') or the part of the sentence contained in the parentheses that immediately follow the quantifier phrase. For example, in ' $\exists x (Fx \wedge Gx) \wedge Hx$ ,' the  $x$ s in ' $Fx$ ' and ' $Gx$ ' are in the scope of the quantifier. The  $x$  in ' $Hx$ ' is not



## 22 Preparatory Background

phrase 'Gx' is not. This makes the reference of the second 'x' confusing. Is this x that is G supposed to be the same as the x that is F? This isn't clear. To fix this, we may introduce parentheses:

$$\exists x (Fx \wedge Gx).$$

Now all variables in the sentence lie within the scope of the quantifier '∃' and we can understand this sentence to be saying: "There is something that is both F and G." To say that a variable lies within the scope of a quantifier is to say that it is a **bound variable**. When symbolizing complete sentences in predicate logic, it is important that all variables be bound by quantifiers.

In general, when one makes an existentially quantified claim, one is saying there exists some thing that is a certain way. What kind of thing we have in mind generally depends on the context. The way logicians put it, this depends on the **domain of quantification**, the set of entities over which the quantifiers range. For example, suppose we used 'Bx' to symbolize 'x is blessed,' and then see the sentence:

$$\exists x Bx$$

What this sentence is supposed to represent depends on the relevant domain of quantification. The domain of quantification may be:

- the set of all entities that exist whatsoever, so that the sentence may be read as: "Something is blessed,"
- the set of persons there are, so that the sentence may be read as: "Someone is blessed,"
- the set of persons in a particular community under discussion, for example, those in this house. Then '∃xBx' would mean: "Someone in this house is blessed."

The relevant domain of quantification is fixed by the context. In later chapters we will see philosophers sometimes making reference to this fact. They will explicitly exploit the fact that our quantifiers may sometimes be restricted, so that they range over a limited set of objects. Or at other times, a philosopher will exploit the fact that in some cases our quantifiers may be "wide open," meaning they range over the largest domain of quantification possible, including any entities whatsoever.

We can now distinguish three types of letter symbols that are used in first-order predicate logic.<sup>6</sup>

- Predicates, which are symbolized by upper case letters: F, G, H, ...
- Names, which are symbolized using lower case letters from the beginning of the alphabet: a, b, c, ...

**Bound variable:** a variable that is within the scope of some quantifier

**Domain of quantification:** the set of objects over which the quantifiers range in a given context, the set of possible values the variables can take

- Variables, which are symbolized using lower case letters from the end of the alphabet: x, y, z, w, u, v, ...

In addition to the existential quantifier, there is also another quantifier, the **universal quantifier**, which is used to symbolize claims involving 'all' or 'every.' For example,

Everyone is happy,  
may be symbolized as:

$$\forall x Hx.$$

We may read this as:

- For all x, x is happy.
- Every x is happy.
- Everyone is happy.

(Note that if the only kinds of entities we ordinarily take to have emotional states like happiness are persons, the relevant domain of quantification is the set of all persons.)

To take another example,

Everyone is a happy philosopher,  
may be symbolized as:

$$\forall x (Hx \wedge Px)$$

or,

every x is such that it is happy and a philosopher.

How would we symbolize 'All philosophers are happy'? This says something different than saying that everyone whatsoever is both happy and a philosopher ( $\forall x(Hx \wedge Px)$ ). 'All philosophers are happy,' is symbolized using the symbol ' $\supset$ ' for 'if ... then':

$$\forall x (Px \supset Hx).$$

We can read this back into English as 'For all x, *if* x is a philosopher, *then* x is happy.' This says the same thing as our original 'All philosophers are happy,' which of course is different than saying 'Some philosophers are happy,' which is expressed in first-order logic as:

$$\exists x (Px \wedge Hx),$$

or 'There exists an x such that x is a philosopher and x is happy.'

**Universal quantifier:**  $\forall$ , a symbol of predicate logic. When combined with a variable, it can be used to represent a statement to the effect that everything is a certain way

**EXERCISE 0.8****Symbolizing Sentences in First-Order Predicate Logic**

Using the key below, symbolize the following sentences in first-order predicate logic.

Key:

a: Alex

b: Barney

Cx: x is clever

Sx: x is a student

Tx: x is a teacher

Rxy: x respects y

1. Alex is a student.
2. Alex is a clever student.
3. Someone is a student.
4. Someone is a clever student.
5. Alex respects Barney.
6. Alex respects someone.
7. Someone respects Barney.
8. Some teachers respect some students.
9. Everyone is a teacher.
10. Everyone is a clever teacher.
11. All teachers are clever.

In the next chapter and throughout the book, we will find that the formulation of theses and arguments in the language of first-order predicate logic is often essential. Particularly when we are considering issues of existence, we will be required to formulate statements in predicate logic. Only then can we be clear about what follows from them. To do so, we will need to have under our belts some basic rules of inference involving existentially and universally quantified statements.

There are four basic rules which are summarized in Table 0.3. Some of these rules are a bit complicated, but for our purposes in this book, the rules that will be used most often are Existential Quantifier Introduction (EI) and Universal Quantifier Elimination (UE). So let's briefly consider some examples using these rules of inference.

Table 0.3 Four Rules of Predicate Logic

<b>Existential Quantifier Introduction (EI)</b>	
From anything of the form:	$Fa$
One may infer:	$\exists xFx$
<b>Existential Quantifier Elimination (EE)</b>	
If it has been established that:	$\exists xFx$
Then one can introduce a new term 'a' into the language whose denotation is not presumed to be identical to any individual already in the domain of quantification and infer:	$Fa$
<b>Universal Quantifier Introduction (UI)</b>	
If one has introduced a new term 'a' as an arbitrary name, and shown for it that:	$Fa$
Then, one may infer:	$\forall xFx$
<b>Universal Quantifier Elimination (UE)</b>	
From anything of the form:	$\forall xFx$
One may infer using any name 'a':	$Fa$

Here is an example of the kind of inference that will be deployed in Chapter 2. Suppose one believes the following:

Humility is a virtue.

This may be symbolized in first-order predicate logic as:

$\forall h.$

Using the rule EI then, we can conclude:

$\exists xVx.$

This may be read back as: There exists some  $x$  such that  $x$  is a virtue.

To consider another example, if we have reason to believe the following:

Plato is a philosopher who taught Aristotle.

We may symbolize this as:

$Pp \wedge Tpa$  (Plato is a philosopher and Plato taught Aristotle.)

And then using EI, we can infer:

$\exists x (Px \wedge Txa)$

In both cases of the application of EI, what we are doing is introducing a variable  $x$  to stand in for a particular subject.

Note then that when you have established an existentially quantified sentence, you may infer that there is something in the relevant domain of quantification that has the relevant features. Thus, from existentially quantified sentences we can often infer that something exists in the domain of quantification that has the relevant features. There is some

## 26 Preparatory Background

$x$  that can stand in as the value of this variable that makes the sentence true. For the sentence,

$$\exists x (Px \wedge Txa),$$

the object that can stand in as the value of the variable to make the sentence true is (as we just saw) Plato.

But, it is worth noting that not all sentences containing existential quantifiers will allow us to infer that there is something that exists that has certain features. In general, even if a sentence contains an existential quantifier, if the quantifier is not the main connective of the sentence, then one is not licensed to conclude that there exists anything with the relevant characteristics. For example, consider these sentences in first-order logic:

$$Fa \supset \exists x Px \quad (\text{read as: If } a \text{ is } F, \text{ then something is } P.)$$

$$\exists x Px \vee \exists x Qx \quad (\text{read as: Either something is } P \text{ or something is } Q.)$$

$$\neg \exists x Px \quad (\text{read as: Nothing is } P.)$$

None of these sentences imply the existence of anything that is a  $P$ . You can tell that immediately because the existential quantifier is not on the outside of the entire sentence.

Finally, we should emphasize a difference between universally quantified and existentially quantified sentences. In general, the way to think about the difference is that existentially quantified sentences tell you that something exists, whereas universally quantified sentences (those with a ' $\forall$ ' on the outside) say that everything is a certain way. Universally quantified sentences on their own don't entail the existence of anything. So, for example, if you see a claim like 'All electrons are negatively charged,' we can write this in first-order logic as:

$$\forall x (Ex \supset Nx).$$

However, this sentence on its own doesn't entail that there are any electrons. It just says, if there are electrons, then they are negatively charged. The following sentence entails the existence of electrons:

$$\exists x Ex.$$

So does:

$$\exists x (Ex \wedge Nx),$$

which symbolizes the statement that there exists at least one electron and it is negatively charged. So, if we are looking for claims that imply the existence of something, our attention should turn to those that are existentially quantified, as opposed to those that are universally quantified.

Universally quantified sentences have other uses. They are especially useful when one wants to state universal principles. Examples of universal principles one finds in metaphysical debates are:

<i>Nominalism</i> :	Everything is concrete.	$\forall xCx$
<i>Idealism</i> :	Everything is an idea in a mind.	$\forall xIx$
<i>Presentism</i> :	Only present objects exist.	$\forall x(\neg Px \supset \neg \exists y(y = x))$
<i>Actualism</i> :	Everything is actual	$\forall xAx$

Once one establishes a universal claim like one of these, one can then use universal elimination (UE) to conclude about particular objects in the domain of quantification that they have the relevant features. For example, idealists usually intend their thesis to be comprehensive, in other words, a claim about the nature of everything whatsoever that exists. This implies that the domain of quantification that is relevant to the idealist is the set of all entities that exist. So if one is an idealist and thus believes that everything that exists (whatsoever) is an idea in a mind, then using universal elimination, one can conclude from:

$\forall xIx$ ,

and the fact that (say) Disneyland exists:

$\exists x (x = d)$  (read as: There exists some  $x$  such that  $x$  is identical with Disneyland.)

that:

$Id$ .

This may be read as: Disneyland is an idea in a mind.

Note that in formulating some of the claims in the last pages, we have made use of the symbol '=' to represent the relation of identity. Identity is another two-placed relation like the teaching relation (symbolized above using 'Txy'). It is a relation that is of special interest to metaphysicians and is particularly useful in formulating metaphysical theses. We will have much more to say about identity beginning in the very next chapter.

## EXERCISE 0.9

### Recognizing Valid Argument Forms in Predicate Logic

In the following examples, state which of the four valid argument forms the arguments instantiate (EI, UI, EE, or UE).

- Everyone is mortal. Therefore, Donald Trump is mortal.
- Some humans have free will. Therefore, Donald Trump has free will.
- Socrates lived in the past. Therefore, there exists something that lived in the past.

## Suggestions for Further Reading

There are many excellent critical thinking and introductory logic textbooks available that develop the material introduced in this chapter further. Some excellent critical thinking texts are Richard Feldman's *Reason and Argument* and Thomas McKay's *Reasons, Explanations, and Decisions: Guidelines for Critical Thinking*. Some excellent introductory logic texts are Merrie Bergmann, James Moor, and Jack Nelson's *The Logic Book* and Gary Hardegree's *Symbolic Logic: A First Course*.

## Notes

- 1 See C.S. Peirce "The Fixation of Belief." We will discuss the role of common sense in metaphysics further in Chapter 1.
- 2 Most college philosophy departments offer courses in Critical Reasoning and Logic that develop this material further.
- 3 Deductive validity is the default notion of validity with which philosophers operate. It is controversial whether there is any genuine sense of validity other than deductive validity; however, I put in the qualifier 'deductive' to explicitly contrast this notion with what is sometimes called 'inductive validity.' An inductively valid argument is one in which the premises do not logically imply the conclusion, but the premises make it reasonable to believe the conclusion in some weaker sense of providing evidence for it. For example, from the premise that the sun has risen every day up until now, we may infer the conclusion that the sun will rise tomorrow. This argument is (one might argue) inductively valid but not deductively valid.
- 4 The method for assessing the validity of arguments introduced in Section 0.2 on validity is what is referred to as a *semantic* method because it is based on the *meanings* of the premises and conclusion. In these final sections, we will be introducing methods for assessing validity *syntactically*, that is, based on the *forms* of the premises and conclusion, independent of their specific meanings.
- 5 Of course, we could alternatively use the semantic method introduced in the validity section (0.2) to show that this is a valid argument.
- 6 We can also now explain why this logic is called *first-order* logic. In the kind of predicate logic we are discussing here, variables are used to range over entities (people, cats, dogs, cell phones, and so on). They may be used to replace names, as when we move from  $Ts$  (Shaq is tall) to  $\exists xTx$  (Someone is tall). In *second-order* logic, variables are also introduced to stand for properties or attributes, ways entities are. They may then replace predicates, for example if we wanted to move from  $Ts$  (Shaq is tall) to  $\exists F Fs$  (Shaq is some way). The status of second-order logic is controversial. And this controversy is directly related to the metaphysical issue over the status of abstract entities like properties or attributes. We will discuss this debate further in Chapter 2. For now, we will just continue to use first-order logic.

# 1 An Introduction to Ontology

## Learning Points

- Introduces ontology, a central subfield of metaphysics
- Presents the Quinean method for determining one's ontological commitments, including the method of paraphrase
- Considers the various types of data that get used in arriving at an ontology
- Discusses the use of Ockham's Razor to decide between ontologies

## 1.1 Ontology: A Central Subfield of Metaphysics

In this chapter, we will introduce one of the most central subfields of metaphysics: ontology. **Ontology** is the study of what there is. In metaphysics, just as in science, one of the main things we want to find out is what kinds of things there are in the world.<sup>1</sup> Although in the various sciences, discussion is usually confined to a particular domain of reality – biology may be concerned with the kinds of living things there are, physics with the subatomic constituents of matter – in metaphysics, we want to know what kinds of things there are in a sense that is more general.

Suppose physics tells us that the basic constituents of matter are leptons and quarks. The metaphysician will then ask: Are there only these physical objects, or are there also other types of entities? For example, are there, in addition to these electrons and quarks, also some nonphysical entities, like minds? Are there also abstract entities like numbers or qualities? And in addition to objects (abstract and concrete), are there other categories of entities – events, processes, spatiotemporal manifolds? All of these are ontological questions, questions about what types of entities exist.

In the mid-twentieth century, philosophers like W.V. Quine (1908–2000), inspired by developments in formal logic, initiated a new method for

**Ontology:** 1. The study of what there is; 2. A particular theory about the types of entities there are



addressing ontological questions. This method has since become standard in metaphysics, and it is the main topic of this chapter.<sup>2</sup> In the following two chapters, we apply this method to two specific debates in metaphysics.

Quine's view is presented in his extremely influential paper from 1948, "On What There Is." In this paper, Quine undertakes two projects. First, he argues that many philosophers before him have been misled in matters of ontology. Metaphysicians of the past have been too quick to believe in all manner of controversial things from abstract entities like numbers and qualities (Virtue, Beauty, the Good) to even nonexistent objects (like Pegasus or the Land of Oz).<sup>3</sup> According to Quine, many of these philosophical errors can be traced to an ignorance of matters of logic. An examination of the logical structure of sentences thus plays a large role in Quine's critique. This negative part of "On What There Is" is followed by a positive part in which Quine develops what he takes to be the correct method to decide which entities one ought to believe in; in other words, how to decide one's **ontological commitments**.

**Ontological commitments:** the types of entities one ought to believe in, given the sentences one accepts

## 1.2 The Puzzle of Nonexistent Objects

Quine begins his paper by criticizing what he takes to be a clearly mistaken ontological view – a view according to which there are nonexistent objects. This is a good place for us to begin as well, for seeing the errors with this view will lead us to have a better handle on what is a good method for settling what types of things one should believe in.

First, let's see why anyone would believe there are nonexistent objects. To see the motivation for this surprising view, consider the following two sentences:

Pegasus does not exist.

Santa Claus does not exist.

Both of these sentences are true. But if a sentence is true, it must at least be meaningful. And if a sentence is meaningful, then each part of the sentence must itself have a meaning. But then from this it follows that the word 'Pegasus' means something and the phrase 'Santa Claus' means something. But what are their meanings? 'Pegasus' and 'Santa Claus' are names (they aren't adjectives or predicates), and so their meaning must involve what they name. So, 'Pegasus' names something: Pegasus. And 'Santa Claus' names something: Santa Claus. Therefore, there is something that is Pegasus. And there is something that is Santa Claus. So, from the plain fact that the sentences we started with are true (the sentences saying these things *do not* exist), we are forced into believing that these things are, and so we look to be ontologically committed to them. In other words, these are entities in which we should believe given the sentences we take to be true.

This argument is traced by Quine all of the way back to Plato (c.428 BC-c.348 BC).<sup>4</sup> It seems that just accepting the claim that something does not exist commits us to its being. As Quine puts it in “On What There Is,” “nonbeing must in some sense be, otherwise what is it that there is not?” (1948, pp. 1-2). What are we talking about when we say that Pegasus and Santa Claus do not exist if not Pegasus and Santa Claus?

TEXT BOX 1.1

The Use/Mention Distinction

You will notice that we have adopted the convention of sometimes placing words or phrases in single quotation marks. It is standard practice in philosophy to adopt this convention. In doing so, we are respecting a distinction noted by Quine: the **use/mention distinction**. This is a distinction between two ways in which a linguistic item (a word, phrase, or sentence) may appear. A sentence may use a linguistic item so that it plays its typical semantic role (naming some object if it is a name, modifying some object if it is an adjective, and so on). Or, a sentence may *mention* the linguistic item, using it to refer to itself. In cases where a linguistic item is being mentioned rather than used, the philosophical convention is to place the relevant word or phrase in single quotes. The following table gives examples in which a word or phrase is being used on the left, and examples in which a word or phrase is being mentioned on the right:

Use	Mention
The book is on the table.	The word ‘book’ has one syllable.
Santa Claus does not exist.	Nothing exists for the name ‘Santa Claus’ to refer to.
Please close the cellar door.	‘Cellar door’ is thought to be the most beautiful sounding phrase in the English language.

**Use/mention distinction:** a distinction between two ways in which a word or phrase may appear in a sentence. A sentence may use the linguistic item so that it plays its typical semantic role (naming some object if it is a name, modifying some object if it is an adjective, and so on). Or, a sentence may mention the linguistic item, using it to refer to itself. In cases where a linguistic item is being mentioned, a philosophical convention is to place the relevant word or phrase in single quotes.

At first appearances, this is an absurd conclusion. How could what is nonexistent in some sense be? Although Quine ultimately rejects the conclusion that there are nonexistent objects, he notes a couple of things that a philosopher could say to make the thesis more intelligible. He conjures two philosophers, whom he names McX and Wyman,<sup>5</sup> in order to consider two positions as to what these nonexistent things might be:

VIEW 1 (McX): Pegasus, Santa Claus, and other nonexistent entities are not concrete objects existing somewhere in the world, but ideas

in the mind. E.g., Pegasus is not a live, physical horse with wings, but the idea of a horse with wings.

VIEW 2 (Wyman): Pegasus, Santa Claus, and other nonexistent entities are "unactualized possibles." They are just like any other entity, except they lack the properties of actuality and existence. E.g., Pegasus is a real, physical horse with wings; just not one that exists in our actual space and time.

Although both of these views may on reflection appear to be natural ways to think about what we are talking about when we talk about nonexistent entities, Quine has objections to both of these views. First, to McX's position, Quine argues it is unsatisfactory to say that Pegasus and Santa Claus are mere ideas in the mind. This is because (recall) the whole issue of nonexistent entities was raised by our initially noting that sentences like 'Pegasus doesn't exist' and 'Santa Claus doesn't exist' are true. But, of course, the *ideas* of Pegasus and Santa Claus do exist. It is rather Pegasus (the flying horse) and Santa Claus (the jolly man carrying presents) who do not. So, the view of McX must be wrong.

Quine voices several objections to Wyman's view, and these are more subtle. First, let's try to understand better what Wyman's view is. Wyman is appealing to a distinction between objects that are actual and those that are merely possible. In our everyday life, we are used to thinking that there are very many things that are possible: phones that take pictures, elevators to the Moon, flying cars, world peace, but only some of what is possible is actual.

According to Wyman, when we say that Pegasus and Santa Claus don't exist, we aren't saying there aren't such entities. Instead, we are only saying that these entities are among the group of merely possible things, things that are possible but are not actual (more like world peace than phones that take pictures). In doing so, Wyman is understanding existence as a property of only actually existing things. There is a wider class of entities that have *being*, that *are*. Those entities that *exist*, on the other hand, are only those that are actual.

Quine's first objection to the view that nonexistent entities are unactualized possibles involves rejecting the distinction between *being* on the one hand and *existence* on the other. As he puts it:

Wyman . . . is one of those philosophers who have united in ruining the good old word 'exist'. . . . We have all been prone to say, in our common-sense usage of 'exist,' that Pegasus does not exist, meaning simply that there is no such entity at all. (Quine 1948, p. 3)

What after all is the difference between existing and being? There does not seem to be any. Anything that is exists. Anything that exists is. Existence isn't a special feature some objects have that others lack. At least, as Quine points out, this is how we speakers of English have tended to use

the word. Conflating existence with actuality, as Wyman does, is to distort the meaning of 'existence.'

This simple point, that being and existence are the same, is one that analytic philosophers since Quine have taken as a postulate. This is also implicit in the way philosophers (both logicians and metaphysicians) treat the existential quantifier,  $\exists$ , of first-order logic (see Preparatory Background, Section 0.7). ' $\exists x$ ' is taken to mean 'there exists an  $x$  such that' or 'there is an  $x$  such that.' These are just alternative ways of saying the same thing. The philosopher Peter van Inwagen has boiled down the contemporary view of existence, deriving from Quine, into the following four theses:<sup>6</sup>

Thesis 1: *Being is not an activity.* Being isn't something things do like dancing or taking a nap.

Thesis 2: *Being is the same as existence.* There is no such thing as something that is, but does not exist, or vice versa.

Thesis 3: *Existence is univocal.* In other words, the concept of existence has the same meaning whatever kind of entity it is applied to. Chairs and numbers may be different types of entities – one concrete, one abstract – but they do not participate in different kinds of existence. What it means to say they exist is the same in both cases. The difference between chairs and numbers is not in the way they exist but rather in their properties, what some may call their "natures." Just as when we say, "This coin is round" and "This cricket field is round," we are using the word 'round' in the same, univocal sense, so when we say, "This chair exists," and "This number exists," we are using the word 'exists' in the same univocal sense.

Thesis 4: *The single sense of being or existence is adequately captured by the existential quantifier of first-order predicate logic.* When we say, "There is a table in this room," or "There exists a table in this room," what we are saying may be adequately symbolized using the notation of first-order logic, specifically the existential quantifier as in:  $\exists x (Tx \wedge Rx)$ .<sup>7</sup>

So, Quine's first objection to Wyman is that he is distorting the meaning of the word 'exists.'

Quine has another critique of Wyman's view.<sup>8</sup> The moral of this critique can be expressed in the form of the famous Quinean slogan: *No entity without identity.* To defend this claim, Quine asks the reader to consider two examples of what we might think are unactualized possibles. Look to the nearest doorway and imagine that there is a fat man in the doorway. Now imagine that there is a bald man in the doorway. Quine asks the following questions:

Take, for instance, the possible fat man in that doorway; and again, the possible bald man in that doorway. Are they the same possible

man, or two possible men? How do we decide? How many possible men are in that doorway? Are there more possible thin ones than fat ones? How many of them are alike? Or would their being alike make them one? Are no *two* possible things alike? . . . Or finally, is the concept of identity simply inapplicable to unactualized possibles? (Quine 1948, p. 4)

What Quine's slogan "No entity without identity" means is that if something exists, that is, if there is such a thing, then there must be objective facts about what it is identical to and facts about what it is not identical to. Consider any entity that obviously exists, for example, the U.S. basketball player Stephen Curry. Because he exists, in other words, there is such a person as Curry, there are facts about what and whom he is identical to and what and whom he is not identical to. Curry *is* identical to the 2015 MVP (most valuable player) for the National Basketball Association. Curry *is not* identical to the current King of Belgium.<sup>9</sup> If it were vague, or if there were no facts about whom Curry was identical to, if he could be identical to the current King of Belgium, there is just no fact of the matter, then according to Quine's doctrine, we should be skeptical of the very matter of his existence.

This is Quine's second critique of the view that nonexistent entities like Pegasus are unactualized possibles. Quine argues we shouldn't believe there are any unactualized possibles because for unactualized possibles (like the possible fat man in the doorway), there are no determinate, precise answers to identity questions. There simply is no fact about whether there are two possible men in the doorway, just one, or a thousand. There is no fact about whether the possible fat man in the doorway is identical to the possible bald man in the doorway or not. And so we should be skeptical that there are any such things in the first place.

#### TEXT BOX 1.2

##### Numerical Identity vs. Qualitative Identity

When we discuss identity in this book, please always bear in mind that metaphysicians have something very particular in mind, what is ordinarily called **numerical identity** or **identity in the strict sense**.

When you see someone claiming in metaphysics for some objects *a* and *b* that *a* is identical to *b* (or  $a = b$ ), what is usually being said is that *a* and *b* are numerically identical. This means that *a* and *b* are the same object. We may have two names here ('*a*' and '*b*') but there is only one object. This is why the relation is called 'numerical identity,' because it is identity in the numerical sense of being one.

**Numerical identity (or identity in the strict sense):** oneness, the sense of '*a* is identical to *b*' meaning that *a* and *b* are the same object, that they are one

In ordinary speech we sometimes use the word 'identity' or say some  $x$  'is identical to' some  $y$  to express that something weaker than strict identity obtains. For example, we may say "These two cars are identical," or "These two dresses are identical." In such cases, we are not saying the dresses or cars are numerically identical. This would be to say that what appears to be two cars is really only one car; what appears to be two dresses is really only one dress. This would be interesting perhaps, but often in ordinary cases when we say such things we are not ascribing numerical identity to the dresses or cars, but a weaker relation that philosophers call **qualitative identity**. Objects  $a$  and  $b$  are qualitatively identical just in the case they share all of the same qualities (the same color, the same size, the same shape, and so on). Philosophers will often say that qualitative identity is not identity (in the strict sense of the term). It is a weaker relation than genuine or numerical identity.

**Qualitative identity:** the sense of ' $a$  is identical to  $b$ ' meaning that  $a$  and  $b$  share qualities (the same color, the same size, and so on)

It is conceivable that there are other viable answers to the question of what nonexistent entities could be, if there are such things. But for many, Quine's critiques of the two most natural positions on the topic - that of McX and Wyman - have been enough to make them skeptical of the position that nonexistent entities should be part of one's ontology.

### EXERCISE 1.1

#### The Argument for Nonexistent Entities

Put the argument for nonexistent entities presented in the second paragraph of this section in numbered premise form. Before seeing Quine's own response to this argument, which premise or premises of this argument do you think one should consider rejecting?

## 1.3 Finding One's Ontological Commitments: Quine's Method

We have seen that Quine rejects the view that there are nonexistent entities like Pegasus and Santa Claus. Nonetheless, he wants to accept that both of these sentences are true:

Pegasus does not exist.

Santa Claus does not exist.

But, according to Quine, just because these sentences are true and just because they mean something, indeed even though the words 'Pegasus' and 'Santa Claus' themselves are meaningful, this does not commit us to believing in such things as Pegasus and Santa Claus. Pegasus and Santa Claus do not exist, and (what is just to say the same thing another way) they lack being. So, there must be some other way in which 'Pegasus' and 'Santa Claus' are meaningful, since they do not name anything. And, however this turns out, it must not commit us to the existence of winged horses or jolly old men who live at the North Pole.

Quine's view is that once we represent these sentences in the language of first-order predicate logic, it will become clear what the real ontological commitments of these claims are. Quine calls this procedure the process of **regimentation**. We want to regiment the statements we take to be true into a language in which their entailments are clear. When we do this, using 'p' to stand in for 'Pegasus' and 's' to stand in for 'Santa Claus,' we see the structures of these sentences are at first pass:

$$\neg \exists x (x = p)$$

$$\neg \exists x (x = s)^{10}$$

These sentences are meaningful. But we can now see that given their structure, they do not entail the existence or being of anything, let alone Pegasus or Santa Claus. Indeed, the presence of the negation sign in the front shows that they explicitly deny the existence of Pegasus or Santa Claus.

#### **Regimentation:**

the procedure of representing statements in symbolic logic to make it as clear as possible what follows from these statements

### **TEXT BOX 1.3**

#### **Semantic Ascent**

There is a very interesting move Quine makes in "On What There Is" that is characteristic of much of twentieth century philosophy. This is what is known as **semantic ascent**. This occurs when, in order to address one kind of philosophical question, a philosopher "ascends to the semantic plane," addressing first a question about the meaning of certain key terms in the original question.

In our case, Quine attacks his original question (Are there nonexistent entities?), an ontological, metaphysical question, by addressing the corresponding semantic question (Does a sentence like 'Pegasus does not exist' mean anything, and if so what does it mean?). In doing so, he "ascends" from the "ontological plane" up to the "semantic plane." The hope is that once the semantic issue is cleared up, the original ontological issue will be cleared up as well.

**Semantic ascent:** when, in order to address a question, a philosopher "ascends to the semantic plane," addressing first a question about the meaning of key terms in the original question

What about the words 'Pegasus' and 'Santa Claus' themselves and the symbols 'p' and 's' we have used in the regimentations? Must they not mean something in order for the whole sentences to be meaningful? Yes, however, one thing Quine emphasizes in "On What There Is" is that we mustn't confuse the demand that words like 'Pegasus' must *mean* something with the demand that they *name* something. Quine proposes the view (following the logician and philosopher Bertrand Russell (1872-1970)) that names are actually disguised descriptions. And so we find the meaning of a name by finding the description it is disguising. For example, Quine considers the view that 'Pegasus' means the same as 'the winged horse that was captured by Bellerophon.'<sup>11</sup> Then, the meaning of 'Pegasus does not exist,' could be expressed as:

$\neg \exists x (x \text{ is the winged horse that was captured by Bellerophon})$

or, using symbols standing in for the predicates 'is winged,' 'is a horse,' and 'was captured by Bellerophon':

$\neg \exists x (((Wx \wedge Hx) \wedge Cx) \wedge \forall y (((Wy \wedge Hy) \wedge Cy) \supset y = x))$

This sentence is meaningful and yet it does not entail the existence of a winged horse. Instead, it explicitly denies that there is such a thing. The sentence is true just in case there is *no* thing that satisfies the definite description.

Quine's positive view about ontological commitment then can be expressed using another famous slogan: *To be is to be the value of a bound variable*. We are only committed to the existence of something when we accept a sentence that quantifies over it. To be precise, we should accept the existence of all and only those entities needed to stand in as the values of bound variables to make these existentially quantified sentences true. To see what this entails, let us consider an example.

#### TEXT BOX 1.4

##### Names as Definite Descriptions

One might be wondering about the right-hand side of the symbolization here, the clause:  $\forall y (((Wy \wedge Hy) \wedge Cy) \supset y = x)$ . There is a reason this clause must be included in the symbolization.

We noted that Quine's strategy for showing that names like 'Pegasus' can mean something even if they lack reference appeals to Russell's theory of names as abbreviated descriptions. But Russell's view is actually more specific than that. It is not just that names are descriptions, but that they are definite descriptions. The following table illustrates the distinction between definite and indefinite descriptions:



Definite description	Indefinite description
'the winged horse captured by Bellerophon'	'a winged horse captured by Bellerophon'
'the jolly man from the North Pole who brings presents to children on Christmas'	'a jolly man from the North Pole who brings presents to children on Christmas'
'the teacher of Alexander the Great who wrote the <i>Nicomachean Ethics</i> '	'a teacher of Alexander the Great who wrote the <i>Nicomachean Ethics</i> '

Russell argued that the ordinary names in our language ('Pegasus,' yes, but also names that have referents like 'Stephen Curry' or 'Elon Musk') are definite descriptions because definite descriptions are like names in that they aim at picking out a unique object.

For example, consider the third set of descriptions in the table above. As it turns out, it is the Greek philosopher Aristotle (384-322 BC) who satisfies the definite description 'the teacher of Alexander the Great who wrote the *Nicomachean Ethics*.' And so we might think that 'Aristotle' means the same as 'the teacher of Alexander the Great who wrote the *Nicomachean Ethics*.' We would not want to say that 'Aristotle' means the same as the indefinite description, 'a teacher of Alexander the Great who wrote the *Nicomachean Ethics*,' because it is possible that many people could satisfy this description. But if many people satisfied the description, we wouldn't then say that the name 'Aristotle' refers to all of them. This would be a case in which the name is defective since it fails to pick out a unique individual. Russell's point is that it is built into the meanings of 'Pegasus' or 'Aristotle' or 'Stephen Curry' that they denote a single individual.

And now finally we can understand why the symbolization of 'the winged horse captured by Bellerophon' includes that final clause:  $\forall y (((Wy \wedge Hy) \wedge Cy) \supset y = x)$ . Consider this phrase:

$$(Wx \wedge Hx) \wedge Cx.$$

What this says when read back into English is:  $x$  is winged and  $x$  is a horse and  $x$  was captured by Bellerophon. This is how one symbolizes an indefinite description.

To symbolize a definite description, to make it clear that we are not just talking about a winged horse captured by Bellerophon, but that we are talking about *the unique* winged horse captured by Bellerophon, we need to add in an explicit clause stating this uniqueness condition:

$$((Wx \wedge Hx) \wedge Cx) \wedge \forall y (((Wy \wedge Hy) \wedge Cy) \supset y = x)$$

What this says when read back into English is:  $x$  is winged and  $x$  is a horse and  $x$  was captured by Bellerophon, and for any  $y$ , if  $y$  is winged and  $y$  is a horse and  $y$  was captured by Bellerophon, then  $y$  is (numerically) identical to  $x$ .

So, we can see what this last clause ensures is that there is no more than one thing that satisfies the description in the first part of the symbolization. And so to symbolize the whole sentence, 'The winged horse captured by Bellerophon does not exist,' we just add our negation sign and the existential quantifier phrase to bind the variable  $x$ . This results in:

$$\neg \exists x (((Wx \wedge Hx) \wedge Cx) \wedge \forall y (((Wy \wedge Hy) \wedge Cy) \supset y = x))$$

One can find further description of Russell's view on definite descriptions in his paper "On Denoting" from 1905.

Say you take the following sentence to be true, perhaps because you take it to be part of your best theory of the world:

Electrons exist.

To see whether this sentence commits you to the existence of anything according to Quine, we need to regiment it: symbolize it in the language of first-order predicate logic. (In this case, the symbolization is easy. For more complex sentences or those that don't reveal their quantificational structure so clearly as this one, the regimentation process will involve more work, but let's just start with this simple case.) When we regiment this sentence using the language of first-order logic, we see it says this:

$$\exists x (x \text{ is an electron})$$

or, using 'Ex' to stand for 'x is an electron':

$$\exists x Ex$$

Now, what does Quine's theory of ontological commitment say? Let's see what Quine says in his own words:

To be assumed as an entity is, purely and simply, to be reckoned as the value of a variable.... The variables of quantification, 'something,' 'nothing,' 'everything,' range over our whole ontology, whatever it may be; and we are convicted of a particular ontological presupposition if, and only if, the alleged presuppositum has to be reckoned among the entities over which our variables range in order to render one of our affirmations true. (Quine 1948, p. 13)

To tell whether something (an electron, proton, Pegasus, Santa Claus, and so on) should be "assumed as an entity," we must first put the sentences we take to be true into quantificational language, into the language of first-order predicate logic. Then we can see what these sentences quantify over.

Returning to the last case, in order for ' $\exists x Ex$ ' to be true, there must be something that exists that can stand in as the value of the bound variable 'x.' This means there must be some entity that satisfies the description 'Ex.' Thus, we can see clearly that this sentence commits us to the existence of (at least) one electron.

Let's apply Quine's method to a slightly more complicated sentence:

Some electrons are bonded to protons.

Say this is a sentence you accept. Symbolizing this sentence in first-order logic, we get:

$$\exists x \exists y ((Ex \wedge Py) \wedge Bxy),$$

where 'Py' stands for 'y is a proton' and 'Bxy' stands for 'x is bonded to y'

Again, to apply Quine's method for determining our ontological commitments, we need to see what needs to exist in order for this sentence to be true. Here, there must be something to stand in as the value of the bound variable 'x' and something to stand in as the value of the bound variable 'y.' This means there must exist some x that is an electron and some y that is a proton and the x must be bonded to the y. So, this sentence ontologically commits us, according to Quine, to at least one electron that is bound to at least one a proton.

We now see the meaning of the slogan: *To be is to be the value of a bound variable*. This slogan succinctly expresses Quine's view that to determine one's ontological commitments, one should express the statements one takes to be true in first-order logic, and then read one's ontological commitments off as those entities needed to stand in as values of the bound variables in order to make those sentences true. We can summarize this method in these three steps:

*Quine's Method for Determining One's Ontological Commitments*

*Step 1:* Decide which sentences you take to be true.

*Step 2:* Regiment the sentences by symbolizing them in the language of first-order predicate logic.

*Step 3:* Commit yourself to all and only those entities needed to stand in as the values of the bound variables in order to make the sentences true.

## **EXERCISE 1.2**

### **Finding Ontological Commitments**

Regiment the following sentences into the language of first-order logic. Determine what Quine would say you would be ontologically committed to, were you to accept the sentences as true.

1. Some donuts have pink sprinkles.

- a. donuts
- b. sprinkles
- c. pinkness
- d. both a and b
- e. all of the above

2. All donuts have pink sprinkles.

- a. donuts
- b. sprinkles

- c. pink sprinkles
  - d. all of the above
  - e. none of the above
3. Some donuts contain holes.
- a. donuts
  - b. holes
  - c. perforated donuts
  - d. both a and c
  - e. all of the above

## 1.4 The Method of Paraphrase

The three steps listed earlier constitute the main parts of the Quinean method. However, there is also a loophole of sorts. What if you take your best theory of the world and, after regimentation, you find that it commits you to some entities you find distasteful for one reason or another, some entities you would prefer not to believe in? According to Quine, are you thereby automatically committed to them? Well, not necessarily.

In “On What There Is,” Quine imagines he is presented with a biological theory which contains as one of its claims the following:

Some zoological species are cross-fertile.

As we will discuss in more detail shortly, Quine is a naturalist. This means he wants to believe what the best scientific theories tell him.<sup>12</sup> And so, if the best biological theory says some species are cross-fertile, this is something he will accept as true. From here, he symbolizes this sentence in first-order logic to find out what sorts of entities it commits him to. We may use ‘Sx’ to stand for ‘x is a zoological species’ and ‘Cx’ to stand for ‘x is cross-fertile’ and this leaves us with:

$\exists x (Sx \wedge Cx)$

When we do this, we can see plainly that the sentence commits Quine to the existence of species. But what are species? Species are clearly not individual animals like lions or tigers. Lions and tigers are members of species, they belong to species, but are not species themselves. Species are abstract entities, e.g., lionhood (*Panthera leo*) or tigerhood (*Panthera tigris*). Perhaps you might think of them as sets, a kind of mathematical object, that contain individual animals as their members. Quine himself is skeptical of the existence of abstract entities.<sup>13</sup> Quine is happy to accept the existence of individual animals like lions and tigers, but not abstract entities like species.

Then what does this mean for his acceptance of the biological statement we just mentioned? Since this sentence, when regimented, reveals itself to require commitment to abstract entities like species, it looks like Quine must make the choice: either accept abstract entities like species into his ontology or reject the biological theory.

Here is where the loophole comes in. Quine says there is a third alternative:

When we say that some zoological species are cross-fertile we are committing ourselves to recognizing as entities the several species themselves, abstract though they are. We remain so committed at least until we devise some way of so paraphrasing the statement as to show that the seeming reference to species on the part of our bound variable was an avoidable manner of speaking. (Quine 1948, p. 13)

Quine gives himself a way to accept the biological theory while at the same time rejecting the existence of abstract species. He allows that in the process of regimentation (Step 2), one may paraphrase the sentence to avoid quantification over entities one does not wish to countenance.

So, instead of symbolizing ‘Some zoological species are cross-fertile’ as:

$$\exists x (Sx \wedge Cx),$$

which, as Quine notes, would commit him to the existence of species, one might instead regiment it in such a way as to only entail the existence of concrete entities like animals. Here is one attempt based on the understanding that what it means for two species to be cross-fertile is for members of those species to be able to mate with each other and produce offspring.

$$\exists x \exists y ((Lx \wedge Ty) \wedge Mxy) \vee \exists x \exists y ((Bx \wedge Ey) \wedge Mxy) \vee \exists x \exists y ((Zx \wedge Cy) \wedge Mxy) \vee \dots$$

where:

Lx: x is a lion	Zx: x is a zebra
Tx: x is a tiger	Cx: x is a cobra
Bx: x is bear	Mxy: x mates with y to produce offspring
Ex: x is an elephant	

Read back into English, this sentence says “Either some lions mate with some tigers and produce offspring, or some bears mate with some elephants and produce offspring, or some zebras mate with some cobras and produce offspring, or ...” where this sentence is continued to include all possible animal pairings.

It is easy to see that according to the rules of first-order logic, this sentence only commits Quine to the existence of concrete things: individual animals like lions, tigers, and cobras. In the chapters that follow, we will see this method of paraphrase deployed quite often by philosophers wishing to debate the existence of certain types of entities. Even if one's first attempt to regiment a sentence reveals it to quantify over some kind of entity, it will sometimes be possible to produce an alternative regimentation (a paraphrase) that evades that ontological commitment.

One might be wondering at this point whether anything at all goes, whether, that is, one can always apply the method of paraphrase to argue away a class of entities one finds distasteful, as Quine wants to do in the case of species. The answer to this question is "No." There are rules here. Notice where the method of paraphrase comes into the Quinean method for determining one's ontological commitments. One has already completed the first three steps:

*Step 1:* Decide which sentences you take to be true.

*Step 2:* Regiment the sentences by symbolizing them in the language of first-order predicate logic.

*Step 3:* Commit yourself to all and only those entities needed to stand in as the values of bound variables to make the sentences true.

Then, after worrying that one does not want to believe in the entities one comes to in Step 3, one goes back to Step 2 and tries to regiment the sentences in another way that doesn't commit one to the problematic entities. But one can't regiment sentences in any way one likes. One must produce a regimentation that has some plausible claim to capturing what was being said by the original sentence one started with. If your best theory of the world involves a claim like, "There are 10-dimensional strings," there are regimentations of this that are plausible and others that are not. For example, it will be plausible to regiment this as:

$\exists x (x \text{ is a string and } x \text{ is 10-dimensional}),$

but not:

$\exists x (x \text{ is an electron and } x \text{ is 3-dimensional}).$

This second sentence may be expressed in English as, "Some three-dimensional electrons exist." This is not what the sentence we started with said. For a paraphrase to be suitable in an ontological debate, it must plausibly convey what the original sentence was intended to convey.

**EXERCISE 1.3****Producing Paraphrases**

The logician Alonzo Church (1903-1995) once gave a parody of Quine's method in which he suggested that all sentences seeming to quantify over women should be paraphrased into statements about their husbands or fathers. In a sentence seeming to quantify over a woman, one would regiment it instead as quantifying over a man's "secondary presence." In such a way, following Quine's suggestion, it would be possible to eliminate ontological commitment to women altogether, a view Church jokingly referred to as 'ontological misogyny.' As Church put it, "the misogynist is led by his dislike and distrust of women to omit them from his ontology. Women are not real, he tells himself, and derives great comfort from the thought, there are no such things."<sup>14</sup>

Let's assume that Church's ontological misogynist wants us to accept the truth of the following sentence:

Some U.S. Senators are women.

What is the paraphrase the misogynist would recommend?

In addition, explain exactly what makes Church's example so absurd, and how Quine would distinguish Church's case from his example involving zoological species. What makes the latter a more acceptable use of the method of paraphrase?

**TEXT BOX 1.5****A Case of Events**

To illustrate Quine's method, we will consider an argument the philosopher Donald Davidson (1917-2003) gave for the existence of events. Davidson argued for events by claiming that it is only by quantifying over events that we arrive at a satisfactory account of the logical structure of sentences about actions.

To see the argument, consider the following sentence:

(1) John danced in the bathroom at midnight.

One natural symbolization of this sentence in first-order predicate logic is:

$Djbm$ ,

using the three-place predicate ' $Dxyz$ ' for 'x danced in location y at time z.'

Using three instances of Existential Introduction (EI), we may deduce from this:

$\exists x\exists y\exists zDxyz$ .

We can thus see the truth of the original sentence as committing one to a person (the dancer), a location and a time.

But several philosophers, for example, Anthony Kenny in *Action, Emotion, and Will*, noticed problems with this symbolization. One problem is that 'John danced in the bathroom at midnight' would seem to entail:

(2) John danced in the bathroom.

But this raises the question of how to symbolize (2) so there is such a logical entailment. It is natural to think the logical form of (2) is just:

Djb,

using the two-place predicate 'Dxy' for 'x danced in y.' But if this is correct, then (1) does not logically entail (2). Another option would be to represent (2) using the same three-place predicate as (1). The sentence would then be seen as involving an implicit, unvoiced reference to a time. The correct symbolization of (2) would then be:

$\exists xDjbx$

This will make it so that (1) entails (2), but then why think the correct predicate to use to represent (1) and (2) is only three-placed? After all, we might have instead started with the sentence:

(3) John danced the tango in the bathroom at midnight.

And then we would have said the correct predicate to use to symbolize these sentences would be the four-placed 'Dxyzw,' for 'x danced the y in z at w.'

Moreover, what of the sentence:

(4) John danced provocatively in the bathroom at midnight.

This too seems to entail (1) and (2), but how could we incorporate the adverb 'provocatively' into our symbolizations to show this?

In his paper "The Logical Form of Action Sentences," Davidson argues that there is a simple way to represent action sentences that does not raise these issues. His solution not only allows us to see the logical relationship between (1), (2), (3), and (4) (that (3) and (4) entail (1) and (2), and (2) entails (1)). It also does not force us to decide arbitrarily how many places our "dancing" predicate should have, and permits a natural representation of adverbs. His solution is to view all action sentences as quantifying over events. On this view, (1) would best be symbolized as:

$\exists x (((Ixj \wedge Dx) \wedge Bx) \wedge Mx)$

where 'Ixj' stands for 'x involved John,' 'Dx' stands for 'x was a dancing,' 'Bx' stands for 'x was in the bathroom,' and 'Mx' stands for x was at midnight. According to the rules of first-order logic, this entails:



$$\exists x ((Ixj \wedge Dx) \wedge Bx),$$

which Davidson would propose as the correct symbolization of (2). And both of these are entailed by:

$$\exists x (((Ixj \wedge Dx) \wedge Tx) \wedge Bx) \wedge Mx),$$

which would be a natural way to represent the logical structure of (3). The sentence (4) may then be represented as:

$$\exists x (((Ixj \wedge Dx) \wedge Px) \wedge Bx) \wedge Mx).$$

The adverb 'provocatively' can now naturally be seen as modifying the dancing, as it should. To solve all of the problems involved in understanding the logical structure of action sentences, we only need to quantify over events.

Following Quine's view that one is ontologically committed to all and only those entities needed to stand in as the values of the bound variables to make one's sentences true, we thus see how a consideration of the logical structure of action sentences may motivate ontological commitment to events. Today, belief in events is common in philosophy. Indeed, in later chapters, we will see the commitment to events being a central part of metaphysicians' views in the philosophy of time and causation.

**Meta-ontology:** the study of what one is doing, or what one should be doing, when one is engaged in an ontological debate

#### TEXT BOX 1.6

### Ontology and Meta-ontology

When we study *meta-ontology*, we are asking questions about the nature of ontology itself. A view in ontology is a view about what exists. A view in meta-ontology is a view about what one is doing or ought to be doing when one is engaged in an ontological debate.

The method of paraphrase is the final component of Quine's theory of ontological commitment. We now have a useful method for determining one's ontological commitments on the basis of the theories we take to be true.

In the next two chapters we will discuss two central issues in ontology concerning the status of abstract entities and the status of material objects. We will see Quine's method in action there.

## 1.5 Ockham's Razor

Now that we have introduced the Quinean method for determining one's ontological commitments, we would like to go ahead and put this method

to work. Step 1 of the method requires us to first decide which sentences we take to be true. But what is the correct way to find the sentences one takes to be true? Metaphysicians have a host of different views on this topic. One common view is that expressed by Quine himself towards the end of his paper:

Our acceptance of an ontology is, I think, similar in principle to our acceptance of a scientific theory, say a system of physics: we adopt, at least insofar as we are reasonable, the simplest conceptual scheme into which the disordered fragments of raw experience can be fitted and arranged. (Quine 1948, p. 17)

According to Quine, what we are looking for in an ontology is similar to what we are looking for in science. We begin by looking for an overall theory of the world – a set of true sentences that can capture a set of data. Just as in the search for good scientific theories, one norm that guides us is the preference for theories that are *simple*. We prefer theories that can state what the world is like using the smallest set of assumptions, positing the fewest number of kinds of entities.

When Quine says that one should start with the “simplest conceptual scheme” and thus we should prefer theories that entail the fewest number of kinds of entities, he is using the philosophical principle known as **Ockham’s Razor**. This principle, named after the medieval English philosopher William of Ockham (c.1287-1347), states that entities should not be multiplied beyond necessity. One’s ontology should be as simple, as parsimonious, as is possible while still explaining everything that needs explaining. Ockham’s Razor is often expressed in Latin as:

*Entia non sunt multiplicanda praeter necessitatem.*

In ontology as well as in science, one should prefer theories that explain one’s data using the fewest kinds of assumed entities.

Quine expresses this preference for what we might call a **sparse** (vs. an **abundant**) **ontology** by saying he has a “taste for desert landscapes.”<sup>15</sup> The preference for parsimonious ontologies is a preference that is independent of Quine’s general method for determining one’s ontological commitments described in the previous sections. One might use that method and instead have a preference for abundant as opposed to sparse ontologies. One might have a preference for the landscape of the rainforest with its lush variety of things over the austere landscape of the desert. Then one would arrive at a different set of sentences in Step 1 of the Quinean methodology, but one could still proceed from there as Quine recommends, regimenting one’s theory of the world into the language of first-order logic.

**Ockham’s Razor:** the principle that one should not multiply one’s ontological commitments beyond necessity

**Sparse ontology:** an ontology that posits a relatively small number of types of entities

**Abundant ontology:** an ontology that posits a relatively large number of types of entities

That said, it should be noted that Quine's preference for ontological parsimony is not an arbitrary preference, like having a taste for coffee rather than tea in the morning. Rather, many philosophers (and scientists as well) have thought that ontological parsimony is relevant to truth: that a simpler theory, a theory positing fewer entities or kinds of entities, is more likely to be true than a theory which posits more. No argument for this conclusion has yet gained widespread acceptance. One natural thought is that the universe is more likely to be simpler rather than complex. But of course this just pushes back the debate to why a simpler universe is more likely.<sup>16</sup> Others have suggested that it is in general more rational to believe in fewer kinds of entities because then there is less for one to be wrong about.

#### TEXT BOX 1.7

#### Fewer Types of Entities vs. Fewer Individual Entities (Token Entities)

Note that in ontology, what we are generally concerned with is how many *kinds* or *types* of entities we posit, rather than the number of *individual* or *token* entities. The difference may be seen by developing Quine's famous claim about desert landscapes. Note that there may be little difference between deserts and rainforests in terms of how many individual objects there are. Think of the vast quantity of grains of sand in the Mojave Desert. There may be as many grains of sand in a desert as individual items in a rainforest. The interesting difference between deserts and rainforests isn't so much the number of individual things there are, but instead the number of *kinds* or *types* of things there are. The climate of a rainforest sustains thousands of different species of plants and animals. This is what we mean by the difference between sparse and abundant ontologies. Sparse ontologies contain fewer types of things than abundant ontologies. The way this is often expressed is by saying that in science and metaphysics, the interest is with **qualitative parsimony**, rather than **quantitative parsimony**.

#### Qualitative parsimony:

the degree to which an ontology includes fewer types of things

#### Quantitative parsimony:

the degree to which an ontology includes a fewer number of things

### 1.6 Where Should Metaphysical Inquiry Begin?

For now, let's just say one agrees with Quine that in the study of ontology, one should aim at the simplest theory of the world, a theory that commits

one to the fewest kinds of entities. The next question is how to discover this simple theory. There are several salient options for where to begin:

- Common or ordinary beliefs in one's community
- Current scientific theory
- Religious texts
- Some combination of the above.

Where should one's ontological, and more broadly metaphysical, inquiry begin?

Many philosophers start with some combination of ordinary beliefs, science, and religion to begin their metaphysical inquiry. But others reject one or more of these sources of potential information, considering them irrelevant to philosophical inquiry.

For example, whether or not you think religious texts have any place in an attempt to discover what there is depends on your religious views. In this book, we will not attempt to enter the fray over religious matters like the existence of God or the possibility of reincarnation, although these are both metaphysical issues. On these matters, we will remain neutral and not much in this book will turn on them.<sup>17</sup>

When it comes to scientific theory, there is a general consensus that current science provides at least one essential input to metaphysical inquiry. The main point of contention concerns how much say science has in metaphysics. Some philosophers, naturalists, argue that the only source of objective knowledge about what the world is like must come from science. Quine defines **naturalism** as "the recognition that it is within science itself, and not in some prior philosophy, that reality is to be identified and described" (1981, p. 21). The naturalist thinks it is wrong to begin metaphysical inquiry by reflecting on what we happen to believe or what any other source tells us, whether it be our community, religion, our family, or just plain common sense. We should base metaphysics on science alone. The goal of ontology is then to state our best current scientific theories as clearly as possible, regiment them in the language of first-order logic, and read off their entailments.

**Naturalism:** the view that it is within science itself that reality is to be identified and described

One currently popular position, **physicalism**, is even more restrictive in what it sees as the starting point for metaphysical inquiry. Physicalism is typically taken to be the view that it is only from physics itself, not from any other scientific, religious, or folk-scientific theory, that we should start an objective inquiry to find out what exists in our world and what it is like. Physics on its own can provide a complete description of (at least the fundamental) types of things that exist in our world and what these things are like. To the extent that other scientific theories like biology or psychology provide some guide as to what our world is like, their claims must be grounded in some way in the claims of physics.<sup>18</sup>

**Physicalism:** the view that physics alone can provide a complete description of the fundamental nature of our world

**Manifest image:** the picture of the world we, as humans, work with in our ordinary lives using the concepts we have acquired and developed in order to explain our perceptual experiences, introspective sense of ourselves, and daily interactions

**Scientific image:** the picture of world science provides

Although naturalism and physicalism at first glance may seem to rule out input from other sources into metaphysical inquiry, this is not entirely the case. According to the naturalist or physicalist, science is the ultimate arbiter of what exists. But this is consistent with the view that in metaphysics, we should also be considering how our lay or pre-scientific pictures of the world may be made consistent with the picture of the world we get from science. In an extremely influential work, “Philosophy and the Scientific Image of Man,” Wilfrid Sellars (1912–1989) described the central aim of philosophy in terms of the project of reconciling what he called the **manifest image** and the **scientific image**. The manifest image is the picture of the world we, as humans, work with in our ordinary lives using the concepts we have acquired and developed in order to explain our perceptual experiences, introspective sense of ourselves, and daily interactions.

The central issue in the philosophy of mind, the mind-body problem, is just this kind of metaphysical issue. The mind-body problem is usually understood today as the question of how the existence of a mental life, the fact that we have thoughts about the world, that we are conscious, can be reconciled with the fact that we are physical beings, made of matter of the sort that physics describes. How could thinking and consciousness arise from mere physical stuff?<sup>19</sup>

Many of the topics we will engage in later chapters also share this structure. Science in general or physics in particular presents us with a certain view. The question is then how this can be reconciled with other things we ordinarily take to be facts: that we have free will, that time passes, that some things that aren’t actual are still possible, or even simply that there are things like ordinary, material objects: tables, chairs, planets, stars, and human beings.

This brings us finally to the question of the place of ordinary or commonly held beliefs in metaphysical inquiry. As we have seen, the naturalist and physicalist think that science is what ultimately gets to decide what exists. And there are some naturalists who take a particularly hard line here arguing that the perhaps naive, pre-scientific beliefs we use in our day-to-day lives have no place whatsoever in metaphysics. As the naturalist philosophers James Ladyman and Don Ross argued in 2007, common sense and pre-philosophical intuitions have been shown again and again to be wrong – about the fundamental makeup of matter, our place in the universe, and so on. They shouldn’t carry any weight at all in an objective inquiry into what exists. And one doesn’t have to be a staunch naturalist like Ladyman and Ross to reject the role of common sense in metaphysical inquiry. In his book *Material Beings*, van Inwagen argues that there isn’t even a body of opinions that we could identify as common sense that could reasonably guide metaphysical inquiry. According to van Inwagen, common sense is what tells you to taste your food before you salt it or to cut the cards before you deal them; it doesn’t have anything to say about the metaphysical makeup of the world (1990, p. 103).

But this viewpoint is not universally shared. Many philosophers think that commonly held beliefs, reports of our intuitions, and other pre-theoretical data do have a place in metaphysical inquiry. Such beliefs are commonly held for a reason, one may think – because they are true. Just as staunchly as naturalists may take contemporary science to be the ultimate arbiter of what there is, other metaphysicians will claim that common opinion holds a central place in metaphysics. Here, for example, is an excerpt from an interview with the metaphysician Kit Fine:

I'm firmly of the opinion that real progress in philosophy can only come from taking common sense seriously. A departure from common sense is usually an indication that a mistake has been made. If you like, common sense is the data of philosophy and a philosopher should no more ignore common sense than a scientist should ignore the results of observation.<sup>20</sup>

There is additional good reason for thinking one cannot ignore common sense altogether in metaphysics. For metaphysics is about issues that matter to us (or it should be!) and if we throw away common sense altogether, we will lose what we cared talking about in the first place. One might argue that common sense is what tells us that time passes, that we have more certain knowledge of the past than the future, that the number two is even and doesn't have a shape or color, that there exist at least some material objects like tables, chairs, and people. The main debate today concerns how large a role such beliefs should have, and their defeasibility, in other words, how much common sense may be overturned by the results of our best scientific theories or philosophical argument, but it is a less widespread position that common sense has no role to play in metaphysics whatsoever.

#### **EXERCISE 1.4**

##### **The Manifest and Scientific Images**

What are two examples in which the manifest image of the world and scientific image of the world seem to contradict each other?

#### **Suggestions for Further Reading**

The main articles discussed above, Quine's "On What There Is" and van Inwagen's "Meta-ontology," spell out the standard Quinean line on determining one's ontological commitments. There are several interesting criticisms of this view. Two notable critiques are William Alston's "Ontological

Commitments" and Richard Routley's "On What There is Not." We will examine further critiques in the next chapter. The 2009 volume *Metametaphysics*, edited by David Chalmers, David Manley, and Ryan Wasserman, contains many articles defending, criticizing, and examining the historical origins of Quine's method. Kris McDaniel's book *The Fragmentation of Being* challenges the standard Quinean view that there are not multiple modes of being or existence. Peter Hylton's book on Quine is also a useful resource.

On naturalism in philosophy, a useful resource is Jaegwon Kim's essay "The American Origins of Philosophical Naturalism." For an interesting recent development of Quinean naturalism, see Penelope Maddy's 2007 book *Second Philosophy*.

## Notes

- 1 In this book, we will most of the time use the word 'world' to mean the entire universe, not just what exists on planet Earth. We will address the meaning of 'world' more fully in the chapter on modality and possible worlds (Chapter 10).
- 2 This is not to say this method is uncontroversial or that no alternatives to it exist. We will discuss alternatives to the Quinean method throughout this text-book. However, since the Quinean method is the most familiar and commonly used method in ontology today, it will be a good starting point for us.
- 3 Here, and throughout most of the text, we will use the words 'thing,' 'object,' and 'entity' interchangeably.
- 4 The source is the discussion of being and nonbeing in Plato's dialogue *The Sophist*, which was thought to be written in approximately 360 BC.
- 5 'McX' is thought to be a veiled reference to the English, idealist philosopher J.M.E. McTaggart (1866-1925). 'Wyman' is thought to be a veiled reference to the Austrian philosopher Alexius Meinong (1853-1920), a philosopher who explicitly held the view that anything we can think of must have some being and that existence is a mere property of an object, a property any object can either have or lack.
- 6 Van Inwagen also formulates a fifth thesis about existence. This is a description of Quine's method of paraphrase, which will be introduced in Section 1.4.
- 7 Van Inwagen presents these theses in two articles: "Meta-ontology" (1998) and "Being, Existence, and Ontological Commitment" (2009).
- 8 There is a third critique as well: that Wyman's view violates the norm of Ockham's Razor (1948, p. 4). We will discuss Ockham's Razor later.
- 9 This is King Philippe, who became king in July 2013.
- 10 The symbol '¬' stands for 'It is not the case that.' '=' stands for the identity relation. See the final two sections (Propositional Logic and First Order Predicate Logic) of the Preparatory Background on Logic.
- 11 Quine actually considers two views about the meaning of names like 'Pegasus' in "On What There Is." One is that mentioned earlier, that names are disguised descriptions. Another alternative he considers is that names are disguised predicates. Here, the meaning of 'Pegasus does not exist,' would be the same as 'There is no thing that pegasizes' or '∃xPx.'
- 12 Another move, compatible with Quine's method, is to reject the scientific theory and say that the claim "Some zoological species are cross-fertile" is false. This would be to adopt an "error theory" about the biological theory. This is not a move Quine himself was attracted to, but it is a coherent alternative. In general, when metaphysicians adopt error theories, and say that claims of what seem to be well-justified theories are really false, because the entities these claims quantify over do not exist, they then have a new challenge. This is to explain why these claims seemed to be true.

- 13 We will say much more about this skepticism and the abstract/concrete distinction in the next chapter.
- 14 In a lecture presented at Harvard University in 1958. Quine was the Edgar Pierce Chair of Philosophy at Harvard from 1956 to 1978.
- 15 "On What There Is," p. 4.
- 16 See the discussion in Alan Baker (2022). See also the interesting distinctions between varieties of parsimony metaphysicians might seek in Sam Cowling (2013).
- 17 A good anthology covering metaphysical issues in the philosophy of religion is Louis Pojman and Michael Rea's *Philosophy of Religion: An Anthology*.
- 18 We discuss the topics of fundamentality and grounding in Chapter 7.
- 19 We will engage the mind-body problem in more detail in Chapter 7.
- 20 [www.3-16am.co.uk/articles/metaphysical-kit](http://www.3-16am.co.uk/articles/metaphysical-kit)



## 2 Abstract Entities

### Learning Points

- Applies the Quinean method to investigate the ontological status of abstract entities
- Introduces the distinction between abstract and concrete entities
- Examines the One Over Many argument for the existence of universals
- Presents truthmaker theory, a rival method for determining one's ontological commitments
- Compares rival positions on the status of properties, including realism, nominalism, and trope theory
- Introduces the indispensability argument and the current debate over the existence of mathematical entities

### 2.1 More Than a Material World?

In the last chapter, we introduced the Quinean ontological method. This method allows us to decide the answers to questions about which kinds of entities exist. It advises us first to determine which sentences we take to be true, then to regiment these sentences into the language of first-order predicate logic, then finally to read off commitment only to those entities needed to stand in as the values of bound variables in order for those sentences to be true.

In this chapter, we will apply this method to a topic that has occupied metaphysicians since the earliest days of western philosophy in ancient Greece. This is the question of the existence of abstract entities. For the purposes of this chapter, let us take for granted the existence of the

concrete, familiar, material objects of our world, those objects like tables and chairs, rivers and mountains, stars and galaxies, persons and other animals, objects that have particular features like shape and size, that occupy spatial locations, objects that we can see, that we can touch or imagine touching.

We will assume for now that these concrete things exist. The question for us is then, in addition to these concrete, material objects, what else exists? For example, consider the mathematical claims you take to be true. Doesn't mathematics tell us that there are such things as prime numbers, such things as even and odd numbers? Doesn't 144 have at least one square root? Aren't there fractions, at least some of which are less than one and greater than zero? If these claims are true, if we believe mathematics is a true theory, then Quine's method seems to entail that we are committed to the existence of numbers. And if we believe numbers exist, then we must believe that there exist at least some abstract objects. The prime numbers, the evens and odds, the fractions, and the rest are certainly not concrete entities. They are not the kinds of things that occupy spatial locations; we can't see or touch them; they don't have features like shape and size, at least not in the straightforward sense that tables and rivers do. Indeed, we can ask the same question about all of the objects of mathematics, the perfect spheres of geometry, the vectors and groups of algebra, the sets of set theory. Do they exist? And if not, are we forced to say that mathematics is simply false?

We will address the ontological status of mathematical entities later in this chapter. First, we will look into a more general argument for the existence of abstract entities, one that applies not only to mathematical entities but to abstract entities of other kinds as well. The sort of abstract entities on which we will focus attention initially are properties. In addition to concrete, material objects like tables and chairs, many have thought there also exist abstract entities that are the *properties* of these tables and chairs: their shapes, sizes, and colors. Or that there exist at least some fundamental physical properties: mass, charge, spin. Others have argued we should believe that there are certain special properties that we value as human beings and that good people ought to strive after, such as humility, honesty, wisdom, and the other virtues. If we believe any of this, and thus become committed to properties like colors, shapes, masses, or virtues, then we seem committed to the thought that our world is made up of more than concrete, material objects. Instead, we must expand our ontology by letting in abstracta (abstract entities) as well. Let's start by trying to get a bit clearer on what we mean by an **abstract** entity.

**Abstract:** a classification of entities, examples include properties and mathematical objects

**TEXT BOX 2.1****Propositions**

One additional category of abstract entity which we will not explicitly discuss much in this chapter is that of propositions. When philosophers and logicians discuss propositions, they distinguish them from the sentences that express them. Propositions are often regarded as the contents of sentences; they are what those sentences mean. And like sentences, they may be true or false. For example, consider these two sentences:

Everyone is in a good mood today.

Tutti sono di buon umore oggi.

These are two different sentences. One is an English sentence; the other is Italian. But they both express the same proposition. We may pick out the proposition expressed using a 'that'-clause. It is the proposition that everyone is in a good mood today.

While sentences may be thought to be concrete entities - they may be identified with concrete strings of letters on a page, or concrete sounds uttered by a speaker - the propositions sentences express are typically viewed as abstract. We will consider the metaphysics of propositions later in Chapter 10.

**2.2 The Abstract/Concrete Distinction**

There are two main ways in which philosophers understand what is meant by an abstract as opposed to a concrete entity. The first we have already seen. Abstract entities are sometimes distinguished by their lack of certain features. For example, while **concrete** entities have observable properties like shapes and sizes, colors and smells, abstract entities like numbers seem to lack these features. In addition, while concrete entities are thought to all have spatial locations, abstract entities do not occupy space. They are rather often thought of as **transcendent** entities, located outside of space, and perhaps time as well. Finally, it is sometimes thought that only concrete entities have causal influence on surrounding objects. Abstract entities lack such causal influence. You can bang your hand on a chair and this will cause you pain or a bruise, but you can't bang your hand on the number three.

This way of classifying entities as concrete or abstract works nicely for the clear cases, the exemplars of concreteness or abstractness. However,

**Concrete:** a classification of entities that are not abstract; examples include material objects like tables, planets, and rocks

**Transcendent:** a transcendent entity is one that is not located in space or time

Table 2.1 Distinguishing Features of Concrete and Abstract Entities

	<i>Tables</i>	<i>Rivers</i>	<i>Stars</i>	<i>Numbers</i>	<i>The virtues and vices</i>
<i>Do they have shapes?</i>	yes	yes	yes	no	no
<i>Do they have sizes?</i>	yes	yes	yes	no	no
<i>Do they have colors?</i>	yes	yes	yes	no	no
<i>Do they occupy space? Do they have spatial locations?</i>	yes	yes	yes	no	no
<i>Are they observable?</i>	yes	yes	yes	no	no
<i>Do they causally interact with anything?</i>	yes	yes	yes	no	no

Table 2.2 Trickier Cases

	Electrons and other elementary particles	Spacetime	Colors	Particular colors of individual objects	Events
<i>Do they have shapes?</i>	no	yes	no	yes	yes
<i>Do they have sizes?</i>	yes	yes	no	yes	yes
<i>Do they have colors?</i>	no	no	yes	yes	no
<i>Do they occupy space? Do they have spatial locations?</i>	yes	no	depends on view about properties	yes	yes
<i>Are they observable?</i>	no	no	yes	yes	yes
<i>Do they causally interact with anything?</i>	yes	yes	depends	depends	yes

in general, although we might say that these features (listed on the left-hand side of Table 2.1) generally are *marks of* concreteness, they are neither necessary nor sufficient conditions for concreteness. To see this, consider the cases in Table 2.2.

These cases all make trouble for this proposed way of distinguishing concrete from abstract entities. For what we find is that for each case, we get a mix of “yes” and “no” answers. Most philosophers tend to think of the first examples, elementary particles and spacetime, as cases of concrete entities. Properties, whether these be general properties like red or green, or particular instances of these properties, like

the redness instantiated by a particular rose or a particular shirt, tend to be thought of as abstract. Events are generally thought to be concrete. But these determinations of what counts as concrete and what counts as abstract can't be traced to the sort of distinguishing features catalogued above. And so it may be best to find another way to make the distinction between abstract and concrete entities.

One alternative is to view abstract entities as those that are *abstractions from* concrete objects. Here is an example. Consider the table nearest to you right now. This is a clear example of a concrete object. Now consider the color of the table by doing the following. Imagine removing all other features of this particular table, its height, its shape, the material of which it is made, and just leaving this one property of the table, its color. This process of stripping away all other features of the object just to leave this one feature (in this case, its color) is what we mean by 'abstraction.' The table's color is an abstraction then from the table itself. We might also, by abstraction, come to other properties of the table: its shape, its height, its smoothness, and so on.

Let's think through just one more example to see how numbers too may in this way be thought of as abstract entities. Geographers tell us that Japan is made up of 6,852 islands. Consider all of these islands. There are five large ones: Honshu, Hokkaido, Kyushu, Shikoku, and Okinawa Island. That leaves 6,847 smaller ones. We may try to think of all of these islands together in all of their complicated detail, or we may abstract away from all of this detail and just consider this one feature, their number: 6,852. When we do this, we abstract away the particular shapes and sizes of all of the islands, their greenness or brownness, their patterns of settlement, their distances from one another. When we do this, we are just left with their number, they are a collection of 6,852 objects. In this way, we may view not just properties like brownness or smoothness, but also numbers, as abstract entities, as abstractions from concrete objects. Concrete objects may then be thought of as objects that are not the result of abstraction from any other object.

**Abstraction:** 1. A psychological process of considering an object while ignoring some of its features; for example, ignoring all other features of a table to just consider its size; 2. The metaphysical relationship of one entity being an abstraction from another, an entity just like the latter except lacking some of its features

There are two ways to understand this process of abstraction: one is psychological and concerns only how we come to know about abstract objects. The psychological process of **abstraction** is the process of considering an object while ignoring some of its features. There is also another, metaphysical way to think of abstraction. In this sense, abstraction is not something we do in our minds, but a metaphysical relation that obtains between a target object and another that lacks some of the former's features (the abstraction). In the remainder of this chapter, this will be our preferred way of making the abstract/concrete distinction. Abstract objects are those that are abstractions from other objects.<sup>1</sup>

**EXERCISE 2.1****Abstract or Concrete?**

Set aside the issue of whether we are justified in believing in the following entities. *If* such things were to exist, would you consider them concrete entities or abstract entities? Explain your answers.

- A. aesthetic features like beauty
- B. the Big Bang
- C. the center of the Earth
- D. dreams
- E. electromagnetic fields
- F. the fictional characters Romeo and Juliet
- G. God

**2.3 Universals and the One Over Many Argument**

Let's now turn our attention to the ontological status of properties: the shapes and colors, masses and charges, virtues and vices. We will return to focus on mathematical entities at the end of this chapter. Our first question is: should we include properties in our ontology? Should we believe not just that there are round tables, but that, in addition, there is an abstract entity: roundness? Should we believe not just that there are electrons, but that, in addition, there is an abstract entity: negative charge?

Properties like roundness and negative charge would be interesting kinds of entities if they existed, not just because they are abstract, but because they may be examples of a kind of entity that is especially puzzling. They seem to be examples of **universals**: entities that are repeatable, capable of being instantiated at multiple locations at once by several different entities. Think about roundness. How many things in the area around you right now are round? The rim of a coffee mug? A doorknob? The button on a lamp or espresso machine? There are many, many round things in any given place at a given time. And if we think that in addition to these round things, there is also an abstract entity, roundness, then this entity will be a universal since it is capable of being instantiated in all of these many different places at once.

To say that a property is **instantiated** is to say that it is had by an entity. Properties that are instantiated at many different locations (universals) are interesting because when they are instantiated, they appear to be wholly present at each of these locations. It isn't that it is a part of roundness that is present here in this coffee mug and another part of roundness that is present there in the doorknob, as I might have part of

**Universal:** a type of entity that is repeatable, that may be instantiated at multiple locations at once by distinct objects

**Instantiation:** the relation between a property and an entity that has that property

my body (my torso) leaning against a chair and another part of my body (my feet) propped up on a desk. When roundness is multiply instantiated at a given time it is *wholly* present in all of these different locations at one and the same time. All of it is there in that mug. And all of it is again there in that doorknob, at the same time.

So, we have another distinction, that between universals (entities that may be multiply instantiated) and **particulars** (entities that may not). As Aristotle says in *De Interpretatione*:

Now of actual things some are universal, others particular. I call universal that which is by its nature predicated of a number of things, and particular that which is not; man, for instance is a universal, Callias a particular.

Plato (c.428 BC–c.348 BC) and Aristotle (384–322 BC) both believed in universals, but they conceived of them in different ways. Plato believed in universal entities called the **Forms**. These were the fundamental entities in Plato's ontology. Some examples of the Platonic Forms are features like Beauty, Justice, and the Good. Plato believed these universals had several interesting features in addition to being repeatable. They are (i) transcendent, in the sense that they exist outside of space and time (as opposed to **immanent**, and located where their instances are). They are (ii) ideal, which means that they are perfections. The Form of Beauty is entirely beautiful, not at all ugly. The Form of Justice is entirely just, not even slightly unjust. This is in contrast to those things in our spatiotemporal realm that may instantiate Beauty or Justice, concrete things like people, works of art, or political institutions. These things are never perfectly beautiful or perfectly just. According to Plato, although people and political institutions may *instantiate* the Forms of Beauty or Justice, they always instantiate a little of the opposite Form as well. Even the most beautiful works of art, the most just political institutions are just a little bit unbeautiful or unjust. Finally, Plato held that the Forms were the kind of entities that (iii) may only be known through the pure intellect, never through sense perception or observation. Although ordinary, perceptible objects may instantiate Beauty or Justice, these are not perceptible features of them, features of them you may see. What we are able to see are only shadows of these Forms. To come to understand and know of a Form itself, perfect as it is, requires the use of the pure intellect.

**Particular:** a type of entity that is not capable of multiple instantiation

**Forms:** the universals that constitute the fundamental entities of Plato's ontology

**Immanent:** an entity that is located in space and time, where and when it is instantiated

## TEXT BOX 2.2

### A Priori vs. A Posteriori Ways of Knowing

When Plato argues that the Forms are known only through the use of the intellect, not through sense perception, he is making

a distinction which philosophers today would call the distinction between **a priori** and **a posteriori** (or **empirical**) ways of knowing. To say a fact is known a priori is to say that it is known in a way that does not involve observation or sense experience. It can be known just by reflection in one's mind (using a method that is *prior to* experience). When you do a proof in a geometry class, for example, you are using a priori methods to deduce certain facts from other facts. When you reflect on the nature of certain concepts or the meanings of certain terms to deduce facts, this also involves the use of a priori methods. For example, if I tell you that my neighbor is a bachelor, then using a purely a priori method, just by reflecting on what it is to be a bachelor, you can come to know that my neighbor is unmarried.

By contrast, to say a fact is known a posteriori or through empirical means is to say that it is learnt through some process of observation or sense experience. For example, there are many ways of coming to know about the color of the Nile River. You can travel to Africa and see it for yourself. You can view a picture of the Nile online or in a travel book. You can have somebody who has visited it describe it to you. All of these are a posteriori ways of knowing. They involve some kind of empirical observation, whether a direct observation of the thing or fact in question or an indirect observation by way of observation of a record or testimony. Either way, the fact is acquired through the senses, through something that is seen or heard (the object itself, a record, a report).

In Plato's view, you may observe things that are beautiful, institutions that are just. But one never observes Beauty or Justice itself. The Forms cannot be seen or touched or heard, they can only be understood through the intellect.

**A priori method:** a way of knowing a fact or proposition that does not involve observation or sensory experience

**A posteriori method:** an empirical way of knowing a fact or proposition, one that involves observation or sensory experience

Aristotle, on the other hand, took universals out of the transcendental realm and viewed them as existing in the concrete objects that instantiate them. This is to say that Aristotle viewed universals as immanent as opposed to transcendent, instantiated in space and time, and located where the objects are that instantiate them. David Armstrong (1926-2014) was a prominent metaphysician who endorsed an Aristotelian theory of properties as universals. According to Armstrong, properties are abstract, universal (repeatable), in space and time (they are wholly located in each of their instances), and knowable by observation or ordinary empirical means. For Aristotelians, we learn about which universals there are in the same ways we learn about the particular entities there are. If one is a naturalist, as many Aristotelians are, one will think one learns about which universals there are by consulting our best scientific theories.



**Realism about**

**universals:** the view that universals exist and they are mind-independent entities

**Platonism:** 1. The view that there are such things as the Platonic Forms; 2. The view that there are such things as abstract, mathematical entities

**Nominalism:** 1. The view that there are no such things as abstract entities; 2. The view that there are no such things as universals; 3. The view that there are no such things as mathematical entities

**Conceptualism:** the view that universals exist, however they are entities that depend on our mind's grasp of them

**One Over Many:** an argument for realism about universals that starts from a premise about some similarities between a group of objects and concludes that there is a universal (a one) that runs through these individual objects (the many)

**Realism about universals** is the view that there are such things as universals. **Platonism** about universals is the more specific view that there are such things as universals and they have some or all of the features Plato thought they had. Realism or Platonism about universals is usually contrasted with **nominalism**, a position that denies the existence of universals.<sup>2</sup> Realism is also historically contrasted with **conceptualism**, an intermediate position between realism/Platonism and nominalism. The conceptualist rejects nominalism and says that universals do exist. However, they claim that universals are entities that depend on our minds' grasp of them, our abstracting them (in the psychological sense) away from concrete objects. Thus, according to the conceptualist, universals exist but are mind-dependent entities. In general, those calling themselves 'realists' about some class of entities will usually insist that the entities with which they are concerned are mind-independent entities; that the entities in question exist and they would exist even if there were never any humans around to think about them. Even for Plato, who thought that the only way one can come to know about the Forms is through the intellect, these entities are robustly mind-independent. The only way we can come to know about and understand the Forms is through the pure operation of the intellect, but the Forms would exist even if there was never anybody around to think about them.

The classic argument found over and over again in the history of philosophy for realism about universals is what is known as the **One Over Many**. It is the overall form of this argument that we will be most interested in evaluating, but let's start by considering one particular example. In "On What There Is," Quine explicitly attacks a particular example of the One Over Many, one he attributes to his character McX:

*One Over Many (McX's version)*

1. There are red houses, red roses, and red sunsets.

Therefore,

2. These houses, roses, and sunsets have something in common: the universal redness.

(Quine 1948, p. 10)

What makes this an example of a One Over Many is that the first premise of the argument is a simple statement about a group of objects that bear some similarity. In this case, we have three kinds of objects (houses, roses, sunsets) that are all red. In addition, for this to be a One Over Many, in the conclusion we infer that there is a universal (a one) that runs through these objects (the many) and is what they have in common. From the fact that several objects, a "many," have some similarity, it is inferred that there is a "one" that runs through the many and accounts for the similarity.

As we have noted, this form of argument is old. One finds it even in Plato's *Republic*, which is believed to have been written in 360 BC. There, in *Republic* Book 10, one finds the inference from the fact that there is a multiplicity of objects similar in some way to the fact that there is a Form (a Platonic universal) accounting for this similarity:

We are in the habit, I take it, of positing a single idea or form in the case of the various multiplicities to which we give the same name. Do you understand?

I do.

In the present case, then, let us take any multiplicity you please; for example, there are many couches and tables.

Of course.

But these utensils imply, I suppose, only two ideas or forms, one of a couch and one of a table. (Plato, *Republic* Book 10, 596a-b)

This use of the inference, where a universal is generated for every term ('couch,' 'table,' but also 'red' and 'round') we use for a multiplicity of objects, will easily generate an enormous quantity of universals. If one had a preference for a parsimonious ontology, one might have concerns. For this reason, some have suggested that the One Over Many be interpreted more narrowly.<sup>3</sup>

The One Over Many, we may assume, always takes roughly the following form:

*Premise:* There exists some  $x$  that is  $F$  and there exists some  $y$  that is  $F$  (where the  $x$  and the  $y$  are distinct entities, i.e.  $x \neq y$ ).

Therefore,

*Conclusion:* There is a universal  $F$ -ness that the  $x$  and the  $y$  both instantiate.

In McX's version of the argument, there are three objects noted to be similar, but strictly speaking, all that is needed are two similar objects to get the argument going.

There are two ways, then, in which one might use this form of argument to generate a theory of universals. The first way, which is more in line with the passage from Plato, is to allow instances of this argument for every case in which we are disposed to apply a common term to each of a group of objects. In other words, we may allow instances of this argument for any general term we may plug in for ' $F$ .' Here are some examples:

*One Over Many (Red)*

1. There are red houses and red roses.

Therefore,

2. There is a universal, redness, that the houses and roses instantiate.

*One Over Many (Chair)*

1. There are two chairs in this room.

Therefore,

2. There is a universal, chairness, that chairs instantiate.

*One Over Many (Mass One Kilogram)*

1. There are flowerpots and handweights that each have mass one kilogram.

Therefore,

2. There is a universal, mass of one kilogram, that these flowerpots and handweights instantiate.

If Plato is willing to countenance universals in any case where there is a multiplicity of objects falling under some common term, all of the above would be legitimate instances of the One Over Many. And what will result is what metaphysicians call an **abundant theory of universals**. An abundant theory of universals is a version of realism that posits a relatively large number of distinct universals. Here we are considering an extreme case in which there is a universal corresponding to any term that is correctly applied to a multiplicity of entities.

However, there is an alternative view. Some may think we should restrict the allowable instances of the One Over Many. For example, Armstrong (1978b) argued that we should restrict instances of the One Over Many to cases in which the objects in the first premise are “genuinely similar” in a certain respect. This will lead to a **sparse theory of universals**.<sup>4</sup> A sparse theory of universals is a version of realism that posits a relatively small number of distinct universals. On an extreme version, and actually this is the kind that Armstrong himself proposes, there are only universals corresponding to the types of entities recognized by our best physical theories. As Armstrong and others have argued, it is only really our most fundamental physical theories that allow us to discern the genuine similarities among objects, the fundamental properties there are in reality. So, for someone with a sparse theory like this, the only legitimate instance of the One Over Many above would probably be the last one, One Over Many (Mass One Kilogram).

More moderate versions of realism are available. For example, a common position is that there are more genuine similarities in nature than just those described by our best, most fundamental physical theories (as extreme sparse theorists believe), but at the same time, not just any term we can apply to a class of things denotes a genuine similarity in nature (as extreme abundant theorists think). For example, one might think that objects may be genuinely similar in virtue of being round – all round things have an objective, common shape. Perhaps one might

**Abundant theory of universals (or properties):**

a version of realism about universals (or properties) according to which there exists a relatively large number of distinct universals (or properties); in one extreme case, a universal (or property) exists corresponding to any term that is applied to a multiplicity of entities

**Sparse theory of universals (or properties):**

a version of realism about universals (or properties) according to which there exists a relatively small number of universals (or properties); in one extreme case, there are only universals corresponding to the types recognized by our most fundamental physical theories

think as well that objects may be genuinely similar in virtue of being lemons – all lemons have a common genetic structure. Unless one had a very permissive, abundant theory, however, most metaphysicians would deny that the phrase ‘in this room’ corresponds to a genuine similarity in nature. All manner of very different things may be in a given room (shape and size permitting) and so One Over Many (In This Room) is not a legitimate instance of the argument.

Today’s metaphysicians hold a variety of views about the range of allowable instances of the One Over Many. In the next section, we consider why one might think that there are no allowable instances of this argument whatsoever. This will also allow us to see what reason there might be to prefer a less abundant theory of universals, aside from Ockham’s Razor.

### EXERCISE 2.2

#### Generating Instances of the One Over Many

Generate an original instance of the One Over Many in which the “F” you use is a term or phrase corresponding to a genuine similarity between objects. Next, think about the universal that is quantified over in the conclusion of this argument. Does this universal seem to have the three features ascribed to universals according to Plato’s theory of Forms? Why or why not?

## 2.4 Applying Quine’s Method

The One Over Many is a traditional way in which metaphysicians have argued for the existence of universals. But in the last chapter, we introduced a rigorous method for determining the entities to which one is ontologically committed. It would be good therefore to consider what someone who wanted to apply that method would have to say about the One Over Many.

Although Quine’s main aim in “On What There Is” is not to deny the existence of universals, only to reveal the proper way to carry out ontological debates, he does discuss the One Over Many in that paper, disputing its validity.

Indeed it is quite easy to see that the One Over Many, as it is usually presented, relies on an invalid argument form. That is to say, in all of the examples given, the conclusion (2) does not follow from the premise (1). The most straightforward way to show this is by regimenting the premise and conclusion of a One Over Many into the language of first-order

predicate logic. Because the last form of the One Over Many we considered above is one all realists about universals we have considered (from the most sparse theorists to the most abundant theorists) will take as legitimate, let's go ahead and regiment that argument.

Let's use the following key:

Mx: x has mass of one kilogram

Px: x is a flowerpot

Wx: x is a handweight

Then premise (1) may be symbolized as:

$$1_R. \exists x(Mx \wedge Px) \wedge \exists y (My \wedge Wy).$$

From this, using the rules of first-order predicate logic, we know we are only committed to the existence of those objects needed to stand in as the values of the bound variables x and y. But these will both be concrete particulars: a flowerpot and a handweight. Once we see the logical form of (1), we can see it does not commit us in addition to the existence of a universal. (2) simply does not follow. Using the additional notation:

Ixy: x instantiates y

m: the universal, mass one kilogram,

a regimentation of (2) would yield:

$$2_R. \exists z (z = m \wedge (\exists x(Px \wedge Ixz) \wedge \exists y(Wy \wedge Iyz))).$$

The rules of first-order logic do not permit us to move from (1<sub>R</sub>) to anything like (2<sub>R</sub>).

Realists like Armstrong have rejected this point, insisting that on a proper understanding of the One Over Many, it is a valid argument. When there is a genuine similarity in nature, there must be some entity that explains this similarity. Although it might indeed be the case that (2) doesn't follow from (1) as a matter of simple predicate logic, (1) expresses a fact that, according to Armstrong, needs explaining. So, (2) is best taken not as following from (1) as a matter of simple predicate logic, but rather as a consequence of a more specifically metaphysical principle about what sorts of facts need explaining. In effect, realists like Armstrong would argue that the statement of the argument above is an enthymeme. (1) does not entail (2) on the basis of logic alone. But there are some tacit premises the realist about universals is assuming, and when we fill these in, we can find an argument that is valid. Let's restate the argument with the tacit premises made explicit:

1. There are flowerpots and handweights that have mass of one kilogram.

\*. If some flowerpots and handweights both have mass one kilogram, then they are genuinely similar in some way.

\*\* . If a group of objects is genuinely similar in some way, then there must be a common entity they all instantiate, a universal, that explains or grounds the similarity.

Therefore,

2. There is a universal, mass of one kilogram, that these flowerpots and handweights instantiate.

The particularly controversial premise here, the one Quine himself would certainly not accept, is the additional premise \*\*. In "On What There Is," Quine makes the following claim talking about the instance of the One Over Many we started with (McX's version):

That the houses and roses and sunsets are all of them red may be taken as ultimate and irreducible, and it may be held that McX is no better off, in point of real explanatory power, for all the occult entities which he posits under such names as 'redness.' (Quine 1948, p. 10)

Let's set aside the jab Quine makes at the defender of universals by calling these abstract entities they believe in 'occult.'<sup>5</sup> Here we see a genuine disagreement between Quine and the realist about universals. The disagreement may be seen to just be over whether or not we should believe in universals. The side who accepts the One Over Many says 'yes' and will be a realist. The side who denies it may say 'no' and remain a nominalist. But what is going on here isn't merely or even primarily a debate about universals. Rather, what we are seeing is more a debate over the correct ontological method.

Armstrong and others who defend arguments like the One Over Many are disagreeing with Quine about something more than just the existence of universals. They are disagreeing with Quine about whether or not the correct method for deciding one's ontological commitments is just regimenting one's statements into the language of first-order predicate logic and seeing what follows. Armstrong will accept that the existence of universals doesn't follow from (1) or more precisely ( $1_R$ ) as a matter of pure logic. Rather, for Armstrong, sentences like (1) commit us to universals because they require metaphysical explanation. They require an account of what there is in nature that explains the similarities between objects. If this is right, then we are not *just* committed to the existence of the objects needed to stand in as the values of bound variables in the sentences we take to be true. When Quine says the fact that a set of objects are all a certain way may be "ultimate" and "irreducible," he wrongly, in Armstrong's view, thinks such metaphysical explanations may be avoided.

This commitment of Armstrong's, to metaphysical explanations involving universals, is the consequence of a general meta-ontological position: **truthmaker theory**. This theory is based around the truthmaker principle:

**Truthmaker theory:**  
the theory that truths have truth makers, some entities or sets of entities that make them true

(TM) Every truth has a truthmaker. In other words, for every truth, there is some entity or entities that make it true.

When a truthmaker theorist looks at a sentence they take to be true, for example, 'Some houses are red,' or 'This flowerpot has mass one kilogram,' they will claim that there must be some entities that make this sentence true. In a case like this, many truthmaker theorists, following a long tradition in philosophy, will say the truthmaker is a complex entity that consists of individual objects instantiating universals.<sup>6</sup> According to Armstrong, truthmakers are states of affairs (Armstrong 1997).<sup>7</sup> Whether all truths require truthmakers and whether truthmakers must be entities made up of particulars and universals are two contentious issues debated today.

**Metaphysical explanation:** an explanation of a fact that says what there is in the world that accounts for that fact's being the case

### TEXT BOX 2.3

#### Metaphysical Explanation

We will discuss metaphysical explanation in more depth in Chapter 7. A related way to discuss metaphysical explanation uses the notion of "ground." A **metaphysical explanation** is an explanation of a fact that says what there is in the world that accounts for that fact's being the case. It tells us what there is in reality or what reality is like, that makes that fact obtain.

### TEXT BOX 2.4

#### Second-Order Predicate Logic

What if we did not take the correct logic to be simply first-order predicate logic, but instead used a different logic, second-order predicate logic? Then, would the inference from (1) to (2) follow as a simple matter of logic?

It seems that it would. Recall that in first-order predicate logic, if one sees something of the form:

Ma,

or:

a is M,

one is allowed to conclude:

$\exists x Mx$ ,

or:

There exists an  $x$  such that  $x$  is  $M$ .

This is the rule of Existential Quantifier Introduction. Just as in first-order predicate logic there are quantifiers that range over the objects in one's ontology, in second-order predicate logic, there are quantifiers that range over the attributes. In second-order predicate logic, if one sees something of the form:

$Ma$ ,

then one may conclude:

$\exists F Fa$ ,

or:

There exists an  $F$  such that  $a$  is  $F$ .

In English, this says that there exists some way that  $a$  is. And if we start with a sentence like

$\exists x Mx$ ,

we may conclude using this rule of Existential Quantifier Introduction in second-order logic,

$\exists F \exists x Fx$ .

There exists some thing that is some way.

Now consider what happens when we apply this proposed rule to the first premise of our argument. Regimented into logical notation, this premise has the form:

$(1_R) \exists x(Mx \wedge Px) \wedge \exists y(My \wedge Wy)$

we may now replace each occurrence of ' $M$ ' with a new predicate variable, ' $F$ ,' and conclude:

$\exists F(\exists x(Fx \wedge Px) \wedge \exists y(Fy \wedge Wy))$

And this appears to be an alternative way to get us to the conclusion we desired, (2). We are here quantifying over a way these objects are. And this way is: having mass one kilogram. So, the One Over Many does appear to go through if we assume second-order logic and the view that quantification over a variable brings with it ontological commitment to something that can stand in as the value of that variable. However, there are several things to say about this.

First, Quine himself was extremely skeptical of the use of second-order logic. Quine believed that second-order logic had problems in itself, as a system of logic. In particular, while systems of first-order logic have the logical property of completeness, second-order logic is incomplete.<sup>a</sup> In addition, Quine worried about the ontological presuppositions of second-order logic. Maybe it is true that there are such things as properties, but this shouldn't follow from matters of logic alone. First-order logic doesn't make specific assumptions about what are the kinds of entities quantified over. Logic gives us a framework for expressing which arguments



are valid, but it is metaphysical investigation that should be used to find out what kinds of things exist. By contrast, with second-order logic, it seems already built into the rules of the logic itself that there are such things as properties or attributes. Quine thinks it is inappropriate that such issues should get decided already as a matter of logic (see Hylton's book *Quine*, pp. 256-257).

Additionally, the interpretation we have been giving to the quantifiers in second-order logic is not the only one available. Logicians do not all agree with Quine that existential quantification in general expresses ontological commitment to an entity capable of standing in as the value of the relevant bound variable. For example, the logician George Boolos (1940-1996) argued for an alternative way of construing higher order quantification, as a device allowing us to refer to (first-order) entities plurally (1984).

Even if we set aside these worries, it is still important to note that those metaphysicians like David Armstrong who used the One Over Many to argue for the existence of universals do not merely think that (2) follows from (1) as a matter of logic. This might be the case if we use second-order logic. But Armstrong thinks that there is a deeper point in the argument, namely that universals are needed to explain certain facts, including the similarities between particulars. Realists do not standardly appeal to second-order logic to argue for their view.

### EXERCISE 2.3

#### Paraphrasing Away Commitment to Universals

As we have seen, a Quinean will not be compelled by the One Over Many argument because their method for determining ontological commitments does not require them to posit entities corresponding to the predicate terms in their language, only those needed to stand in as the values of the bound variables in order to make the sentences they accept true. Some metaphysicians have argued, however, that there are sentences the Quinean ought to accept that do commit them to universals, by their own methodology. Here is an example raised by the metaphysician Frank Jackson:

Red is more similar to pink than it is to blue.

Symbolizing this sentence in the language of first-order logic, we get:

Srpb;

using the following key:

Sxyz: x is more similar to y than it is to z  
 r: the color red  
 p: the color pink  
 b: the color blue.

And this entails, using three instances of existential introduction (EI):

$$\exists x \exists y \exists z Sxyz.$$

Jackson's claim is that if we accept the original sentence, we appear committed by Quine's own lights to these three colors, universals. That is, we are so committed unless we are able to produce an acceptable paraphrase of this sentence that does not commit us to red, pink, and blue.

It seems there are two options here for the Quinean who wants to both defend her ontological method and deny the existence of universals: either (a) produce an alternative regimentation of the sentence, a paraphrase, that plausibly conveys what the original sentence was being used to convey and yet does not entail the existence of colors, or (b) argue to the effect that the original sentence is really false. What would an attempt at a paraphrase look like here? Is this a promising way for the Quinean to defend their position? Or should they instead try to argue that the original sentence is actually false? What could one do to make it convincing that Jackson's sentence is actually false?

## 2.5 Nominalism and Other Options

We have now seen several varieties of realism about universals, a view often motivated by the One Over Many. We have seen how a Quinean may reject the One Over Many. But one then has to ask: If one rejects the existence of universals, then what view are we left with when it comes to the properties we want to ascribe to things?

We have already briefly discussed nominalism, in the form that Quine offers in "On What There Is." Quine rejects the claim that predicates like 'is red' or 'is beautiful' apply to objects in virtue of the instantiation of universals. He simply claims that the fact that these predicates apply in some cases is irreducible, which is to say it doesn't require any metaphysical explanation. This, as it happens, is only one version of the view today called 'nominalism.' It is a version that some have called **ostrich nominalism** since it refuses to answer the question in virtue of what metaphysically objects are similar or appear to instantiate certain properties. Like an ostrich sticking its head in the sand, the nominalist in this case just avoids the question. Probably a more friendly name for this position is **predicate nominalism**. What the view really amounts to is the claim that there are no such things as properties. The term 'nominalism' comes from the Latin 'nomen' for name. There exist only predicates (words), no universals or properties to which they correspond.

Another version of nominalism that is perhaps more popular is what is called **class** or **set nominalism**.<sup>8</sup> Class nominalism has been attractive to philosophers who, for one reason or another, are skeptical of the existence

**Ostrich nominalism:** a version of nominalism that denies the existence of properties and refuses to answer the question of what it is in virtue of which objects are similar or appear to have certain features

**Predicate nominalism:** a view denying the existence of properties; predicates may be satisfied or not satisfied by objects, but there need be no property that exists to explain this fact

**Class (or set) nominalism:** the view that properties are to be identified with the classes (or sets) of objects that instantiate them

of universals (perhaps because they are entities that may be wholly present in many places at once), but are not similarly skeptical of classes and so will appeal to classes to give an account of what properties are.

Classes and sets, unlike universals, are particulars. They are entities postulated by set theory, a branch of mathematics. Sets or classes are entities that have members (except for the empty set,  $\emptyset$ ). They are tools used by logicians to identify the extensions of predicates: the extension of a predicate is the class of objects satisfied by it. For example, the extension of 'is red' is the class of all of the red things. The extension of 'is beautiful' is the class of all of the beautiful things. The view of the class nominalist is that when we use a word like 'red,' we are never referring to a universal, but instead we are referring simply to the extension of the predicate, the class containing all and only the red things. Similarly, 'beauty' refers to the class of all and only the beautiful things. For the class nominalist, properties are classes, the classes of entities that have them.

There are a few things to note about this view. First, if one was skeptical about the existence of universals because they were abstract entities, then one will not be any happier with this version of nominalism. Since properties are now being identified with abstract objects, sets or classes, a class nominalist is still committed to abstract objects. However, one might think, especially if one is convinced by the indispensability argument we will discuss in the next section, that we have more reason to believe in this sort of abstract entity than the universals that are motivated by the One Over Many argument. This is something Quine himself came to believe.

Next, one should note that there is an important objection to class nominalism. This is the **Objection from Coextension**. Isn't it possible that there exist two or more properties that are instantiated by exactly the same group of things? For example, suppose a skilled and dedicated collector fancies a particular shade of blue. Call it 'sky blue.' The collector searches the world and succeeds in collecting all of the sky blue objects there are. He places them in a museum, the SBM (Sky Blue Museum). He paints every surface in the SBM sky blue and then locks the door forever. It will then turn out that the class of objects that are sky blue is exactly the same as the class of objects that are in this museum. That is, the predicates 'is sky blue' and 'is in the SBM' will have exactly the same extension, apply to exactly the same objects. But if class nominalism is true, then properties are identified with their classes, the classes of objects that satisfy them. And then it would follow that 'is sky blue' and 'is in the SBM' refer to the same properties. But these phrases refer to distinct properties. One is a color. The other is a location. So class nominalism must be wrong.<sup>9</sup>

The class nominalist has a response to this objection.<sup>10</sup> She may say that properties are not identical to classes of things that *actually* instantiate a property. Rather they are identical with classes that include both the

**Objection from Coextension:** an argument against class nominalism that there are more properties than those that may be recognized by the class nominalist, since two predicates may have the same extension and yet refer to distinct properties

things that *actually* and the things that *possibly* instantiate the property. It might be the case that actually all of the sky blue objects are in this one museum. However, it could have been the case that they were not, that the collector was unsuccessful or never born in the first place, and so the sky blue objects were scattered all over the world. When we look at properties this way, as classes containing all and only the objects that actually or possibly instantiate them, our two classes differ. For although the two classes will overlap in all of the members that exist at the world where the collector is successful (because all of the actual sky blue things are contained in the SBM), they will have different possible objects as members. There are possible sky blue things that never get placed in the museum; things that it may be impossible to get into the museum. This solution to the problem of coextension is appealing since it captures the intuitive reason we have for thinking that even if all of the sky blue objects are in the SBM, being sky blue isn't the same as being in the SBM - because something *could* have been sky blue and *not* been in the museum.

This response helps with the problem of properties that actually happen to be instantiated by the same class of objects. Though, as many have noted, there is still a challenge accounting for properties that may be shared not just by all actual objects, but by all actual and possible objects. For example, all and only triangles have the following two properties: trilaterality (having exactly three sides) and triangularity (having exactly three angles). This is true not just of all actual triangles. This is true of all possible triangles, whatever triangles there possibly ever could have been. So the class of all actual and possible triangular things is identical to the class of all actual and possible trilateral things. But again it seems these are two distinct properties. And so we should not identify properties with classes. It is possible for the class nominalist at this point to just bite the bullet and claim that these are not really distinct properties after all, if we cannot even imagine them failing to be instantiated together. But the realist about universals will likely not find this response compelling.

A final thing to note about class nominalism - one might wonder if the class nominalist has any way of making a similar distinction between a sparse ontology of properties and an abundant one, as we saw the realist about universals doing above. David Lewis (1941-2001) is one metaphysician who is not particularly moved by the One Over Many argument, but is sympathetic to Armstrong's view that there is a preferred rather small class of attributes that make for the objective resemblances of objects in our world. Or, another way of seeing the motivation for a sparse ontology of properties: Lewis is sympathetic to the idea that there may be a small number of basic properties in virtue of which it is possible to explain the behavior of all things at this world. One who accepts a sparse theory of universals, like Armstrong, can say that it is in virtue of this small class of universals (perhaps just fundamental physical universals

like masses, charges, and spins) that all other features of our world can be explained. Our question now is: if one rejects the existence of universals altogether, is there another way to have a sparse theory of properties? Remember, properties for the class nominalist are not entities objects instantiate in virtue of which they are similar. Rather they are just abstract classes of objects. The main reason to think class nominalism can only be an abundant theory, one with lots and lots of properties and not just a small number, is because set theory yields the existence of classes corresponding to all of the pluralities of objects there are in the world.

Lewis has argued that it is important to have a sparse account of properties. But one does not need a theory of universals to do it. Appealing to a small basic number of universals as Armstrong does is one option, but another option is to simply believe that certain classes of objects are distinguished in some way from other classes of objects. Lewis calls this feature that distinguishes certain classes from others 'naturalness.' The classes/properties that are **natural** are those in which each member is similar to every other in some unique way. In addition, it is in terms of these natural properties (natural classes), that one can give a complete account of the behavior of all things in our world. Most of the properties we talk about every day, properties like being a chair or being a student or being famous, are not natural properties. These are not properties whose members are all objectively similar in some one way. These are not properties that will appear in our final scientific theory of the universe, the theory that has the resources to explain everything. But, if one is inspired by the need to account for similarities between objects, and wants to make a distinction between those properties that make for genuine resemblance between objects and those properties that do not, Lewis has shown how the class nominalist can do it: by appealing to a distinction between natural and non-natural properties.

In addition to class nominalism, there is another position on the nature of properties that also rejects the existence of universals. Like realism, this view grants the existence of mind-independent properties as abstract entities that are more than classes of concrete particular objects. But like nominalism, this view denies that there are any universals – any entities that may be instantiated in multiple locations at the same time. This position is known as **trope theory** or the **theory of abstract particulars**. It is a position that is most often associated with the philosopher D.C. Williams (1899–1983). The trope theory says that properties like shapes and sizes are not, as Plato and Aristotle thought, entities capable of multiple instantiation. Rather each object that instantiates a property instantiates its own particular color or own particular shape. These properties are abstract (you may think of them as abstractions from more concrete entities as we discussed earlier) and yet they are particular

**Natural properties:** a term introduced by David Lewis to distinguish those classes of objects in which each member is exactly and objectively similar to each other in some unique way

**Trope theory (or the theory of abstract particulars):** the theory that properties are tropes, or abstract particulars

(they are not capable of being multiply instantiated). Another word for these abstract particulars is **trope**.

There are several appealing features of a trope theory as opposed to a theory of universals or class nominalism. One attraction is if you thought that there was something problematic about the very idea of a universal, about the idea of something that could have a kind of presence through instantiation in many places at once, then the trope theory gives one a way to avoid this commitment while still accepting the existence of properties. Second, Williams argues that we have more reason to believe in the existence of tropes than in the existence of universals since tropes are what we encounter first in the world. What we observe most directly are tropes, this particular redness here, that blueness there, not the corresponding universals. In addition, trope theory allows one a more parsimonious ontology than one that has two categories, one for particulars and one for universals, and so it should be preferred to realism about universals for this reason.

To see this last point, we should note that the way trope theory is usually defended is not just as the claim that there are such things as tropes, abstract particular properties, or even as the claim that the only kinds of properties there are are tropes. Rather, trope theory is usually defended as the more ambitious claim that the only kinds of things there are whatsoever are tropes. Ultimately everything there is is just a trope or a more or less complicated relation between tropes (another trope). Tropes are the “alphabet of being.”<sup>11</sup>

For example, trope theorists argue that what we ordinarily regard as concrete particulars like tables, chairs, and people, are best regarded as collections of tropes. A table is just a collection of tropes, a particular shape, size, color, texture, and so on. In this way, we needn't believe in two categories, universals and particulars that instantiate them, but just one category of being: abstract particulars. Williams suggests that the trope theorist can build up universals too. Universals are complicated collections of tropes. The universal redness, for example, will be composed of all of the red tropes. Whether what results is a genuine universal in the sense the realist about universals intends is questionable. However, in this way at least the trope theorist can secure referents for terms like 'redness' or 'roundness.'

**Trope:** an abstract particular, e.g., the shape of the Empire State Building

## EXERCISE 2.4

### Four Theories of Properties

We have now seen four theories of properties: realism about universals (in its several forms), predicate nominalism, class nominalism, and trope theory. Which view of the four seems most attractive to you? In a paragraph, explain this theory's appeal.

## 2.6 Mathematical Objects

We have focused so far on a particular kind of abstract object, properties. However, one might think that numbers and other mathematical objects constitute a distinct kind of abstract entity, one that there is independent reason for taking with ontological seriousness.

Plato regarded numbers as kinds of universals, Forms, alongside the other kinds of universals: Beauty, Wisdom, and the rest. And one can easily formulate a version of the One Over Many that will get us to a conclusion about the existence of numbers. For example,

1. There are two moons of Mars, two houses of the British parliament, and two lions guarding the New York Public Library.

Therefore,

2. These moons, houses, and lions have something in common: the universal two (or duality).

Again, Quine will use the resources of first-order predicate logic in order to show how (1) does not commit us to the existence of a number two. It only commits us to moons, houses, and lions (lion statues). And the defender of universals may use meta-ontological considerations about explanation or truthmaking in order to defend their use of this argument. In this section, we will consider a quite different way to argue for the existence of numbers and other mathematical objects, a way that has been more influential over the past decades. This is via the **indispensability argument**.

The indispensability argument is historically associated with Quine and Hilary Putnam (1926–2016). It is an argument for realism about mathematical entities. Realism about mathematical entities is very often called Platonism in the philosophy of mathematics literature. We may state the indispensability argument in the following form with two premises:

1. We ought to have ontological commitment to all that is indispensable to our best scientific theories.
2. Mathematical entities are indispensable to our best scientific theories.

Therefore,

3. We ought to have ontological commitment to mathematical entities.<sup>12</sup>

The first premise follows from a commitment to naturalism. A naturalist, recall, is someone who believes that one should use our best scientific theories in order to settle questions of ontology. The second premise is a substantive claim about our best scientific theories. It is not a philosophical point per se. It is a claim about the content of these theories, that

**Indispensability argument:** an argument for realism (Platonism) about mathematical entities from the premises that (1) we should be committed to all the entities that are indispensable to our best scientific theories, and (2) the claim that mathematical entities are indispensable to our best scientific theories

they make essential reference to mathematical entities. Mathematics is a pervasive element across the sciences, from fundamental physics to the other natural sciences, such as chemistry, biology, and neuroscience, and even today the social sciences, such as economics and political science. Indeed, one might even think, and this is what it means to say mathematical entities are *indispensable* to our best scientific theories, that these theories could not even be accurately stated without reference to mathematical entities. Quantification over mathematical entities cannot be paraphrased away without doing damage to the explanatory success of these scientific theories.

What is especially interesting about the indispensability argument is that it gives one a way to see numbers and classes, indeed all of the mathematical tools used by our best scientific theories, as on a par with other entities that are far less controversial. Nominalists (those who deny the existence of abstracta, including mathematical entities) sometimes motivate their position by noting that we cannot observe mathematical entities and so lack reason to believe in them. But what the defender of the indispensability argument can note is neither do we observe many of the entities posited by our best scientific theories – we can't see or touch electrons or magnetic fields, arguably, we don't observe stock markets or recessions either – but still we are justified in adding them to our ontology because of the role they play in these extremely successful theories (physics, economics).

One influential way of responding to the argument has been to question its first premise. The problem is not so much with naturalism as a general doctrine about using science to guide metaphysics. Instead, the issue is with the statement that one should have ontological commitment to *all* that is indispensable to science. One might argue that some kinds of representations used in science are not intended to have ontological import. In interpreting our scientific theories in order to draw out their ontological implications, contra Quine, we shouldn't blithely adopt commitment to everything the theories quantify over.

Penelope Maddy uses an analogy with the atomic theory of matter in the mid-nineteenth century to motivate this point. At this time, the physics community had substantial justification for the atomic theory of matter, though many remained skeptical of the existence of atoms. As she elaborates:

Although atomic theory was well-confirmed by almost any philosopher's standard as early as 1860, some scientists remained skeptical until the turn of the century – when certain ingenious experiments provided so-called “direct verification” – and even the supporters of atoms felt this early skepticism to be scientifically justified. This is not to say that the skeptics necessarily recommended the removal



of atoms from, say, chemical theory; they did, however, hold that only the directly verifiable consequences of atomic theory should be believed, whatever the explanatory power or the fruitfulness or the systematic advantages of thinking in terms of atoms.... If we remain true to our naturalistic principles, we must allow a distinction to be drawn between the parts of a theory that are true and the parts that are merely useful. (Maddy 1992, pp. 280-281)

Maddy is not suggesting that the metaphysician hold off altogether in drawing metaphysical conclusions from scientific theories. Eventually the physics community as a whole came to endorse the existence of atoms, and we should follow their lead. Indeed, Maddy notes (see the last sentence in the passage above) that even when the physics community was reluctant to endorse the existence of atoms, they still had no problem endorsing other metaphysical ramifications of the atomic theory. So the issue is, even if we are naturalists, should we endorse the first premise of the indispensability argument? Scientists themselves do not always take all indispensable elements of their best theories to have ontological significance. (Surely the atoms were an indispensable element of the atomic theory of matter.) Maddy's point is that naturalism itself seems incompatible with according ontological significance to all parts of scientific theories, as the first premise of the indispensability argument assumes.

Another sort of case many have used to challenge the first premise of the indispensability argument is the widespread use of idealizations in science. **Idealizations** are false assumptions introduced into a theory in order to make it simpler to use. For example, in thermodynamics, we find the ideal gas law (which assumes the existence of particles that do not interact), and in mechanics, physicists routinely make the assumption that certain surfaces are frictionless. But no one believes the particles in a gas fail to interact. And no one believes there are such things as frictionless surfaces. These are false assumptions made to make problem-solving simpler, laws easier to state. Idealizations are rampant not only in physics, but throughout the natural and social sciences. This suggests that the way to draw ontological conclusions from our scientific theories may be more complicated than the first premise of the indispensability argument suggests.

Another way one might respond to this argument involves challenging the second premise that mathematical representations are indispensable to our best scientific theories. To challenge the argument in this way, one would have to provide good reason to think that one could still express the claims of our best scientific theories without any mathematics. This is a project that Hartry Field undertook for one part of physics in his 1980 book *Science Without Numbers*. Field's project was to reformulate Newtonian gravitational physics in a way that didn't quantify over

**Idealization:** a false assumption introduced into a theory in order to make it simpler to use

any mathematical entities. Field wanted to show that a lot of what physicists wish to express using mathematics could be expressed using reference to spacetime instead. Field's project is thus an example of Quine's method (including the use of paraphrase) discussed in Chapter 1. Field's work is brave and interesting, however whether he succeeded in showing that reference to mathematical entities is dispensable is controversial. To show that mathematics is not indispensable to science, one would have to extend the project beyond Newtonian gravitational physics to the rest of science, or show why in principle there is good reason to think Field's method will so extend. This is something that has so far not been shown, but what Field did accomplish gives some nominalists hope.

Taking a step back from the indispensability argument, we will close this chapter by considering a more general problem that mathematics is thought to raise for metaphysics. This problem was expressed by the philosopher Paul Benacerraf in his influential paper "Mathematical Truth." Benacerraf argued that in deciding the ontological implications for mathematics, we are pulled in two directions by two competing desires. Take any mathematical truth, for example, that two is an even number. First, one would like to have a semantics for this mathematical truth, an account of what this proposition means and what makes it true. In addition, one would also like to have a plausible **epistemology** for this truth, an account that explains how we know it. But, Benacerraf argued, when one tries to achieve an attractive semantics for mathematical truths, one is forced to accept a mathematical ontology that yields a very unattractive epistemology. On the other hand, if one tries to achieve a plausible epistemology for mathematical truths, then one is unable to achieve a plausible **semantic theory**. There is thus a choice one has to make, and this choice directly relates to the question of realism (Platonism) versus nominalism about mathematical entities. This is what is known as Benacerraf's dilemma. Like any **dilemma**, we are faced with a choice between two options, each of which yields some unattractive consequences.

Let's first try to understand the challenge that results when one tries to give a plausible semantic theory for a mathematical truth. Benacerraf asks us to consider the following two sentences:

- (1) There are at least three large cities older than New York.
- (2) There are at least three perfect numbers greater than 17.<sup>13</sup>

It looks as if these sentences share the same logical form and that they are both true. But what makes these sentences true? One way to find their truth conditions is to regiment them in first-order logic and read off from there what needs to be the case if these sentences are going to be true. When we do this, it looks like both commit us to the existence of certain kinds of entities.

**Epistemology:** the theory of knowledge and justification

**Semantic theory:** an account of a proposition's or a set of propositions' meaning and truth conditions

**Dilemma:** a choice between two options, each of which yields unattractive consequences

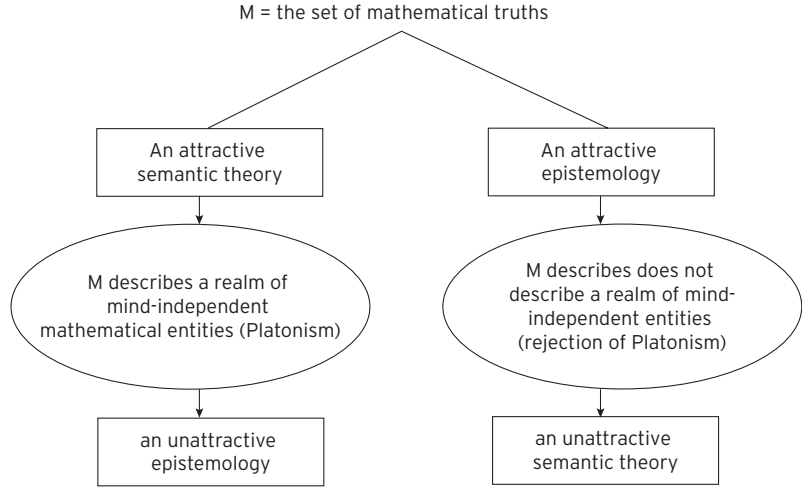


Figure 2.1 Benacerraf's Dilemma

$$(1_R) \exists x \exists y \exists z (((Lx \wedge Cx) \wedge Oxn) \wedge (((Ly \wedge Cy) \wedge Oyn) \wedge (((Lz \wedge Cz) \wedge Ozn) \wedge ((x \neq y \wedge y \neq z) \wedge x \neq z)))$$

$$(2_R) \exists x \exists y \exists z (((Px \wedge Nx) \wedge Gxs) \wedge (((Py \wedge Ny) \wedge Gys) \wedge (((Pz \wedge Nz) \wedge Gzs) \wedge ((x \neq y \wedge y \neq z) \wedge x \neq z)))$$

In the first case, Quine's method tells us we are committed to the existence of at least three cities. In the second case, we are committed to the existence of at least three numbers. Benacerraf's conclusion is, if we want to fit mathematical truths like (2) into a plausible semantic theory, one we already use to understand a wide range of claims, then we are forced to be Platonists, realists about mathematical entities.

Having a plausible semantics for mathematical truths would be nice, but Benacerraf argues that since this will involve endorsing Platonism, the consequence is that one cannot then have a plausible epistemology for mathematical truths. For in general, the way we come to know about the existence of objects is by their causal interaction with us. We know there are more than three cities older than New York because we have visited, been told about, or read about these cities. All of these ways of gaining knowledge involve causal interactions, some less direct than others. But if Platonism is true, then mathematical sentences quantify over mathematical objects. And if there are such things as mathematical objects, then presumably these are not objects located in space, objects we can see and otherwise causally interact with. If such things exist, they are, as Plato thought, transcendent entities. An epistemology of mathematics, a theory of how we come to know the truths of mathematics, then cannot depend on our causal interactions with them. We are left then with a question of what a plausible epistemology of mathematics could be.

Benacerraf notes that there is one promising way to provide an epistemology for mathematical truths. One might think that mathematical truths are known because we prove them. Mathematical knowledge doesn't come from our being causally connected somehow with a set of objects like the natural numbers or algebraic groups, but instead by our fitting certain sentences into a system containing a basic set of axioms or postulates and deriving other sentences using the system's methods of proof. Viewing the epistemology of mathematics this way, we are not disposed to view mathematics as a set of sentences about some class of mind-independent objects, but rather as a set of sentences that have their truth conditions determined just by facts about what follows from what deductively. But if this is correct, one cannot take a view about the semantics of mathematical statements that is similar to the semantics we use to understand the truth of other ordinary statements. To return to the example above, we will not understand the truth conditions of (2) as very similar at all to those of (1), because (2) will not be a sentence that refers to a class of three or more objects. It will instead be about what can be proven in a formal system. This is a problem since it forces us to give up the standard semantic theory for the case of mathematical truths. We are forced to adopt a nonunified semantic theory for our language, one theory for mathematical claims, a distinct one for other claims. Moreover, the resulting account of in virtue of what mathematical propositions are true seems implausible as a semantic theory. That there are at least three perfect numbers is known perhaps because we can prove it, but the fact that we can prove it is not what makes it true. Proof is one thing; truth is something else.<sup>14</sup>

### EXERCISE 2.5

#### The Indispensability Argument

State an example from a scientific theory that would help support the indispensability argument for mathematical Platonism. What is the best strategy for the nominalist to use to respond to this example?

### Suggestions for Further Reading

Sam Cowling's *Abstract Entities* is an up-to-date overview of the relevant ontological issues. On universals and nominalism, David Armstrong has written several influential texts including the two-volume, *Universals and Scientific Realism*, and the shorter *Universals: An Opinionated Introduction*. Several classic texts on the nature and existence of properties can be

found in the volume edited by D.H. Mellor and Alex Oliver called *Properties*. On the distinction between natural and non-natural properties, see David Lewis's *On the Plurality of Worlds* and his influential article, "New Work for a Theory of Universals." For a helpful discussion of nominalism, its motivations and the arguments against the position, see Zoltan Szabó's paper, "Nominalism."

On the indispensability argument in the philosophy of mathematics, the classic statement may be found in Hilary Putnam's *Philosophy of Logic*. One may also consult Mark Colyvan's book, *The Indispensability of Mathematics*. Alan Baker's "Are There Genuine Mathematical Explanations of Physical Phenomena?" provides some good examples of mathematics as it is used in science to support the indispensability argument. For a selection of classic texts in philosophy of mathematics including the debate about Platonism, the collection *Philosophy of Mathematics*, edited by Paul Benacerraf and Hilary Putnam, is recommended.

## Notes

- 1 This definition also allows us to make room for abstracta that are abstractions from other abstract objects. For example, primeness is an abstract feature of the number three. The property of being a shape is an abstract feature of the property roundness.
- 2 As we will see at the end of this chapter, metaphysicians also use these terms to denote similar views about the existence of mathematical entities.
- 3 Plato scholars (for example, see Julia Annas, *An Introduction to Plato's Republic*, Chapter 9) note that elsewhere (even in the *Republic*), Plato doesn't seem to think there is a Form corresponding to every multiplicity whatsoever. Instead, the number of different Forms is more restricted.
- 4 The sparse/abundant distinction is due to the metaphysician David Lewis who introduced it in his *On the Plurality of Worlds*. See also Jonathan Schaffer's "Two Conceptions of Sparse Properties."
- 5 What do you think Quine means by this?
- 6 See for example, Bertrand Russell's *The Problems of Philosophy*, Chapter 11.
- 7 A sidenote: as we know, redness is not one of the universals recognized by Armstrong's sparse ontology. And so it won't be the instantiation of a universal redness that makes 'Some houses are red' true. But in this case, there will be some more fundamental features of the house (or the particles that compose it) that do correspond to universals and that explain the truth of this sentence. Ultimately it might turn out that the truthmaker for this sentence is a very complex state of affairs involving many elementary particles instantiating a few fundamental physical universals. The important point for us is that Armstrong will argue the truth of the sentence requires the instantiation of universals by particulars.
- a To say a logical system is complete is to say that every valid argument expressible in that system is able to be proved in that system.
- 8 For the purposes of this chapter, we will use 'set' and 'class' interchangeably. Both are abstract entities introduced in the part of mathematics called 'set theory.' One feature distinguishing classes from sets that has made the former particularly attractive to some mathematicians and logicians, is that classes cannot be members of other classes. For an accessible introduction to the basics of set theory geared to philosophy students, see the first chapter of David Papineau's *Philosophical Devices*.

- 9 D.C. Williams discusses another example in "On the Elements of Being." It looks like the class of featherless bipeds is the same as the class of things that have a sense of humor. But even if the extensions of 'is a featherless biped' and 'has a sense of humor' are the same, this does not mean we should identify the properties.
- 10 You will find this response in David Lewis's *On the Plurality of Worlds*, Chapter 1. We will discuss this response more in Chapter 10.
- 11 One trope theorist who rejects the claim that everything there is reduces to tropes is John Heil (2012).
- 12 This is the version presented in Mark Colyvan's book *The Indispensability of Mathematics*.
- 13 A perfect number is a positive number that is equal to the sum of its positive divisors. For example 6 ( $= 1 + 2 + 3$ ) and 28 ( $= 1 + 2 + 4 + 7 + 14$ ).
- 14 Or so some would think. There is one theory in the foundations of mathematics, intuitionism, that would say otherwise.

# 3 Material Objects

## Learning Points

- Evaluates the notion of a material object
- Introduces Leibniz's law and some related principles about identity
- Presents the paradoxes of material constitution, the Problem of the Many, and the Special Composition Question
- Examines the problem of metaphysical vagueness

## 3.1 What is a 'Material' Object?

In Chapter 2, we asked whether in addition to all of the concrete material objects (objects like tables and chairs, rivers and mountains, stars and galaxies, persons and other animals), we should also believe in the existence of abstract objects: universals, sets, or numbers. In this chapter, we will take a step back and ask about these material objects themselves.

What do we mean by 'concrete material object'? By 'object' we just mean something that exists. Although it is difficult to give a perfectly satisfactory definition of 'concrete,' we saw in Chapter 2 that one may distinguish the concrete objects as those that cannot be understood as abstractions from something else. It is now time to consider what is meant by 'material.'

One common understanding of 'material' is that which occupies space and persists through time. René Descartes (1596-1650), in the seventeenth century, defined material substances as those that are extended in space. Today, we tend to allow that something may be material even if it strictly speaking lacks spatial extension but rather is point-sized. This includes the fundamental particles of physics, such as electrons and quarks. This is why we say what is material "occupies space" as opposed to "is spatially extended." When it comes to the question of persistence over time, we

will discuss in a later chapter the possibility of time slices of individuals, temporal parts of objects that have zero extension in time. Given the possibility of material objects like this that exist for only a moment, we might also require of material objects only that they have location in time, not that they necessarily persist for longer than a moment.

In addition to these spatiotemporal criteria, it is possible that there are other features that determine what makes an object material. Isaac Newton (1643-1727) defined mass as the quantity of matter in an object. This suggests another criterion: that material objects are those that possess some non-zero mass. There *are* fundamental particles introduced by physics such as the photon that have zero mass. One might think that these provide counterexamples to mass as a criterion for materiality. (It used to be thought that neutrinos were massless as well, but this has now been rejected.) However, physics makes a distinction between those particles that are material, or constitute the matter of the world, and those that instead govern the interactions, and are force carriers. The photon is a force carrier particle; it is what carries the electromagnetic force. So we may hold onto this part of our characterization. Let's then say for the purposes of this chapter that material objects are those that are located in space and time (though they might not have extension in space or duration in time) and have mass.

The metaphysician Jessica Wilson has argued that these scientific criteria should also be supplemented with an a priori, metaphysical one: a "No Fundamental Mentality" requirement.<sup>1</sup> That way, we can ensure that if something is a material object, it either lacks mentality, or any mental features it has are explained by some more fundamental ones. Whether we endorse this type of requirement or not, we should recognize that what it is to be a material object is at least partly an empirical matter and something we may revise as our physical theories change.<sup>2</sup>

Many metaphysicians today are realists about at least some concrete material objects.<sup>3</sup> The question is: which ones to accept? A common position is that there are at least those objects described by the Standard Model of contemporary particle physics: leptons (a category including electrons and neutrinos) and quarks. After that, there is much disagreement about what else exists; in particular, whether in addition to these basic concrete objects there exists anything else, any complex objects composed out of the basic ones. (There is also disagreement over whether in addition to the concrete *material* objects there are also immaterial concrete objects like souls or minds, a topic we will return to in Chapter 7.) Many philosophers hold the view that there exist concrete, material objects corresponding to any way there is of composing simple, material objects (fundamental particles) into complex ones. But in contemporary metaphysics, we also find ingenious arguments for other views. This includes the surprising view, defended by Peter van Inwagen, that the only concrete objects that



exist are fundamental particles and organisms. According to van Inwagen, there exist electrons, flowers, and human beings, but no tables or chairs, mountains or rocks! Even more radically, we find philosophers (for example, Peter Unger) arguing that not only do mountains and rocks fail to exist, but there are no human beings either. Indeed, one of Unger's most famous papers bears the title, "I do not exist."

We will start this chapter by thinking about some classic paradoxes concerning concrete material objects. We will then discuss two issues that arise when we assume a fundamental level of basic material particles: the Problem of the Many and the Special Composition Question. Both have generated a lot of discussion among metaphysicians over the past several decades and show us that the existence of complex material objects like ourselves isn't a trivial matter.

### 3.2 The Paradoxes of Material Constitution

In this section we will discuss two paradoxes involving material objects. Let's set aside the issue of whether or not there is a fundamental layer of basic concrete material objects, and just focus on the ones with which we are most familiar from everyday life: artifacts (tables and chairs, ships and computers), organisms (human beings, cats, trees), and other medium-to-large-sized, natural, inanimate objects (rocks, mountains, planets).

Perhaps you have already heard of the Ship of Theseus. The story goes: Theseus had a large wooden ship that he sailed from Crete to Athens. After some time, the ship needed repairs as its planks started to rot. Gradually the Athenians replaced the planks of Theseus's ship with new planks. After many years, all of the wood of the original ship was replaced with new planks. By this time, the ship contained not a single plank of the original wood. At this point already, an interesting metaphysical puzzle arises about whether what has survived is the original Ship of Theseus. However, the philosopher Thomas Hobbes (1588-1679) added an additional wrinkle

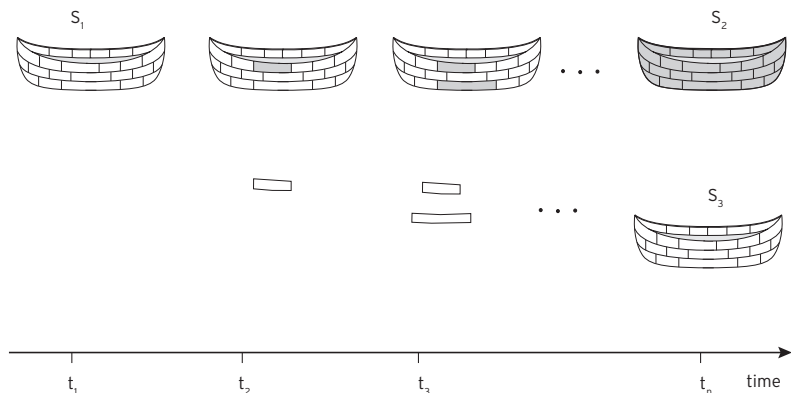


Figure 3.1 The Ship of Theseus

to the puzzle. Suppose the original planks were not destroyed. Instead, as each was replaced, the original planks were stored and finally used to assemble another ship. Soon, two ships stood side by side.

Let's call the original Ship of Theseus that arrived in Athens from Crete ' $S_1$ .' Without assuming whether it is or is not identical to  $S_1$ , call the ship that resulted from the gradual replacement of rotten planks with new planks ' $S_2$ .' And call the ship that was assembled from the old, rotten planks ' $S_3$ ' (Figure 3.1).

Now we can ask: which of the two ships resting next to each other on the shore of Athens is the original Ship of Theseus? We seem to have four options.

- Option 1:*  $S_1 = S_2$  but  $S_1 \neq S_3$  (Only the repaired ship is identical to the original ship.)
- Option 2:*  $S_1 = S_3$  but  $S_1 \neq S_2$  (Only the ship constructed of the original planks is identical to the original ship.)
- Option 3:*  $S_1 = S_2$  and  $S_1 = S_3$  (Both ships are identical to the original Ship of Theseus.)
- Option 4:*  $S_1 \neq S_2$  and  $S_1 \neq S_3$  (Neither ship is identical to the original ship.)

We will look at each option and see what there is to be said for it. But first let's just dwell for a moment on the use of the identity sign '='. When this symbol is used to make a metaphysical claim, it is important that one always reads it as saying "is identical to." In mathematics, one uses the same symbol to express equality. There you will read '=' as 'equals' as in: 'Three plus two equals five.' In metaphysics, '=' refers to the metaphysical relation of numerical identity.<sup>4</sup> When we say " $a = b$ ," this expresses the claim that  $a$  and  $b$  are the same entity. If this identity claim is true, then even though we may have two names ' $a$ ' and ' $b$ ,' these two names are referring only to one thing. For example:

Mark Twain = Samuel Clemens  
 Jay-Z = Shawn Carter  
 Mt. Everest = Sagarmāthā

These statements all express true identity claims in the strict sense we are concerned with. When we say there is an identity, we are not just asserting the entities are similar in some way (their qualitative identity). Rather we are saying they are literally the same thing. There are not two people, Jay-Z on the one hand and Shawn Carter on the other. There is only one person to whom we may refer using either name.

So, let us return to the case of the Ship of Theseus. The first answer to our question, which ship is identical to the original ship, is that it is only  $S_2$ , the ship that resulted from the gradual replacement of parts. Here are some thoughts that motivate this answer. First, think about what happens after the first plank on the original ship has become rotten and gets

replaced. Call the resulting ship, which is just like the original ship except for the replacement of one plank: ' $S_1^*$ .' Surely, we want to say that  $S_1^*$  is identical to  $S_1$ , that it is the same as the original ship. After all, the only difference between  $S_1$  and  $S_1^*$  is one plank. Surely  $S_1$  can undergo a small repair and maintain its identity. If you've ever owned a car, you've probably at some point had to replace a tire or a fender. One doesn't usually think that making such a repair causes your car to cease to exist and a new, numerically distinct car to suddenly appear. Rather, we usually think the car after the repair is identical to the car before the repair. It's the same car; it just has a new part. If we grant this, then we can imagine a succession of ships,  $S_1^*$ ,  $S_1^{**}$ ,  $S_1^{***}$ , and so on, where each ship in the sequence differs from the one before it only by one plank. And just as we may say that as you make small repairs to your car, it doesn't cease to be your car, but maintains its identity over time, so we can say that Theseus's ship doesn't cease to be Theseus's ship as small repairs are made to it, but it maintains *its* identity over time. We can thus conclude of the sequence leading from  $S_1$  to  $S_2$  (the ship that results from the final small repair) that the identity relation holds at each step:

$$S_1 = S_1^* = S_1^{**} = S_1^{***} = \dots S_1^{***} \dots^* = S_2$$

### TEXT BOX 3.1

#### Equivalence Relations

The identity relation has three important formal features:

*Reflexivity*: the identity relation is reflexive. In other words,  
 $\forall x \ x = x$ .

*Symmetry*: the identity relation is symmetric. In other words,  
 $\forall x \forall y \ (x = y \supset y = x)$ .

*Transitivity*: the identity relation is transitive. In other words,  
 $\forall x \forall y \forall z \ ((x = y \wedge y = z) \supset x = z)$ .

Any relation that has these three features (reflexivity, symmetry, and transitivity) is what logicians call an **equivalence relation**. Identity is one equivalence relation. Others include:

- being the same size as
- being the same shape as
- is a member of the same family as.

From here, it is just a short step to get us to the claim that  $S_1 = S_2$ . To establish this, we only have to note that the identity relation is transitive. In other words,  $\forall x \forall y \forall z ((x = y \wedge y = z) \supset x = z)$ .

Since there is a chain of identities linking  $S_1$  to  $S_2$ , we may conclude that  $S_1 = S_2$  by the following argument:

1. The replacement of one plank does not destroy a ship.
  2.  $S_1 = S_1^*$  (From premise (1))
  3.  $S_1^* = S_1^{**}$  (From premise (1))
  4.  $S_1 = S_1^{**}$  (from premises (2) and (3) and the transitivity of identity)
  5.  $S_1^{**} = S_1^{***}$  (from premise (1))
  6.  $S_1 = S_1^{***}$  (from premises (4) and (5) and the transitivity of identity)
  - ...
  - ...
  - n-2.  $S_1 = S_1^{***...}$  (from premises (n-4) and (n-3) and the transitivity of identity)
  - n-1.  $S_1^{***...} = S_2$  (from premise (1))
- Therefore,
- n.  $S_1 = S_2$  (from premises (n-2) and (n-1) and the transitivity of identity)

This is the reasoning motivating the claim that  $S_1 = S_2$ .

This argument is compelling, but there is also reason to say that the original ship is identical to the ship built of the original materials,  $S_3$ . The simple reason is that  $S_1$  and  $S_3$  are made of the same constituents; not just qualitatively similar constituents, planks that look the same, but numerically identical constituents, the very same planks. If we adopt a plausible-looking principle that says "Same constituents, same object," then we will be forced to say that  $S_1 = S_3$ . Moreover, one might think that  $S_3$  is just the result of taking  $S_1$  apart and putting it back together again. If you think of things that way, then it can seem very plausible that  $S_1 = S_3$ .<sup>5</sup>

Given that there is something to be said both in favor of the claim that  $S_1 = S_2$  and that  $S_1 = S_3$ , one might think that perhaps Option 3 is the best response to the puzzle of the Ship of Theseus. There is a good motivation for thinking that  $S_1$  and  $S_2$  are identical, due to the continuity one finds through the gradual transition from  $S_1$  and  $S_2$ , which just amounts to a replacement of parts. There is also a good motivation for thinking that  $S_1$  and  $S_3$  are identical, since they are composed of identical parts. We might thus conclude that both identity statements are true. However, this claim

is problematic. Indeed, it is probably the most problematic of the four options. The reason is that, as we have noted (see Text Box 3.1), identity is an equivalence relation. And this means that if  $S_1 = S_2$  and  $S_1 = S_3$ , then, by the symmetry and transitivity of identity,  $S_2 = S_3$ . Here is the argument:

1.  $S_1 = S_2$  (initial assumption)
2.  $S_1 = S_3$  (initial assumption)
3.  $S_2 = S_1$  (from premise (1) and the symmetry of identity)

Therefore,

4.  $S_2 = S_3$  (from premises (2) and (3) and the transitivity of identity)

But  $S_2$  is not identical to  $S_3$ . Remember what it would mean to say that  $S_2 = S_3$ . This would be to say that there are not two ships,  $S_2$  and  $S_3$ , but just one. But of course  $S_2$  and  $S_3$  are distinct ships. They are sitting at different locations. They are made of different materials. Therefore, they are not just one ship, but two. And so, we cannot accept Option 3.

There is actually an important metaphysical principle that is tacitly being appealed to in this last point. This is the principle of **Leibniz's law**, also known as the **Indiscernibility of Identicals**. It is named after the philosopher and mathematician Gottfried Leibniz (1646-1716). Leibniz's law says that if there is something that has a property that something else lacks, then these things are not identical. If they were the *same* thing, then they would have to share all of the same properties. The very same thing cannot both have and lack a property. (This would be a contradiction.) More precisely, we may state Leibniz's law in either of the following two ways (using the resources of second-order predicate logic):

$$\begin{aligned} \forall x \forall y \forall F ((Fx \wedge \neg Fy) \supset x \neq y) \\ \forall x \forall y \forall F (x = y \supset (Fx \equiv Fy)) \end{aligned}$$

Sometimes you will see Leibniz's law (and other metaphysical principles) expressed using the box symbol '□' attached to it:

$$\begin{aligned} \Box \forall x \forall y \forall F ((Fx \wedge \neg Fy) \supset x \neq y) \\ \Box \forall x \forall y \forall F (x = y \supset (Fx \equiv Fy)) \end{aligned}$$

This is because metaphysical principles like Leibniz's law are often taken to express facts not just about what contingently happens, but rather necessary truths about what things must be like no matter how the world happens to turn out.<sup>6</sup> The box should be read aloud as 'Necessarily.'

**Leibniz's law (or the Indiscernibility of Identicals):** the principle that necessarily, if a and b are identical, then they must share all of the same properties

**TEXT BOX 3.2****The Indiscernibility of Identicals and the Identity of Indiscernibles**

Leibniz's law or the Indiscernability of Identicals should not be confused with a more controversial principle that is also associated with Leibniz: the **Identity of Indiscernibles**. The Identity of Indiscernibles says that necessarily, if *a* and *b* are qualitative duplicates (i.e. qualitatively identical), then they are the same object (i.e. numerically identical). This principle is expressed by the logical formulas:

$$\Box \forall x \forall y \forall F ((Fx \equiv Fy) \supset x = y)$$

$$\Box \forall x \forall y (x \neq y \supset \exists F (Fx \wedge \neg Fy))$$

Necessarily, if objects share the same features, then they are identical. And necessarily, if any objects are not identical, then there must be some feature that one object has that the other lacks.

Leibniz's law is not controversial. For most metaphysicians, Leibniz's law just expresses some basic consequences of what is meant by numerical identity. However, when a philosopher endorses the Identity of Indiscernibles, they are making a substantive metaphysical claim: that there could not be two distinct objects that share all of the same features.

The philosopher Max Black (1909-1988) once proposed the following counterexample to the Identity of Indiscernibles:

Isn't it logically possible that the universe should have contained nothing but two exactly similar spheres? We might suppose that each was made of chemically pure iron, had a diameter of one mile, that they had the same temperature, colour, and so on, and that nothing else existed. Then every quality and relational characteristic of the one would also be a property of the other. Now if what I am describing is logically possible, it is not impossible for two things to have all their properties in common. (Black 1952, p. 156)

The case of Black's spheres certainly seems to involve no contradiction, and thus it would seem to be logically possible. Since the Identity of Indiscernibles is supposed to express a necessary truth about identity, the logical possibility of these spheres would seem to show that the principle is false ... or at least much more controversial than its converse, Leibniz's law!

**Identity of**

**Indiscernibles:** a principle stating that necessarily, if *a* and *b* are qualitative duplicates, then they are identical

**EXERCISE 3.1****Leibniz's Law**

A. Suppose someone presented the following argument for the conclusion that Superman is not identical to Clark Kent:

Superman and Clark Kent cannot be the same person because Clark Kent wears glasses and Superman does not. Also Superman is faster than a speeding bullet and Clark Kent is not. Therefore, by Leibniz's law, Superman and Clark Kent are not identical. They must be different people.

How could one respond to this argument? Does the case of Superman and Clark Kent violate Leibniz's law?

B. Suppose someone presented the following argument for the claim that Angelina Jolie and Russell Brand are the same person (are identical):

Angelina Jolie and Russell Brand are the same person. They were both born on June 4, 1975. They both have long brown hair. They are both movie stars. They have both been married at least twice. Therefore, by Leibniz's law, they must be identical.

How would you respond to this argument?

Applying Leibniz's law back to the case at hand, we may be sure that  $S_2 \neq S_3$  because  $S_2$  and  $S_3$  differ with respect to many properties. They differ with respect to their materials, their location, and so on. And so, since they are discernible (they have different properties), they cannot be identical.

Now, one might try another move and, noting that neither  $S_2$  nor  $S_3$  has a better claim to being the original ship, instead say that both  $S_2$  and  $S_3$  are new ships. This is another possibility. It is Option 4. Here, we may allow there are important similarities between the ship that originally arrived in Athens and the ships that exist now, but say that strictly speaking, the original ship no longer exists. It has been replaced with two new ships.

This last option is appealing since it does not force us to choose between  $S_2$  and  $S_3$  to say which has a better claim to being  $S_1$ . And it presents us with a more plausible option than Option 3, which required us to say that what appeared to be two ships sitting next to each other was really just one ship. However, there is still something unsatisfying about this last option in that it forces us to deny that either ship is the original Ship of Theseus. And, given that there doesn't seem to be any other

plausible candidate, we are forced to say that the Ship of Theseus has now ceased to exist.<sup>7</sup> What's more, recall that we originally had a very compelling argument for the claim that  $S_1 = S_2$ . If nobody had collected the original planks and assembled  $S_3$ , then there would be no threat to claiming that  $S_1 = S_2$ . But how could facts about the identity of  $S_2$  be affected by a process that happened somewhere else?

This is why the Ship of Theseus presents a puzzle to metaphysicians. One has to choose between Options 1, 2, and 4. Does one want to bite the bullet and accept that the Ship of Theseus no longer exists today and there are only two new ships,  $S_2$  and  $S_3$ ? Or is there some way to undermine the motivation we have presented for saying either that  $S_2$  or  $S_3$  is a good candidate for being the original Ship of Theseus?

### EXERCISE 3.2

#### The Ship of Theseus

What is your preferred solution to this puzzle: Option 1, 2, or 4? Defend it against the remaining options.

A second classic puzzle about concrete, material objects is the Puzzle of the Statue and the Clay. A sculptor takes a lump of clay (call this object 'Lump') and makes a statue of the warrior Goliath (call this object 'Goliath'). Suppose that at noon, there is only the lump of clay, but at midnight, the statue Goliath is finished and placed alone on a pedestal. Here is the question: at midnight, should we say that there are two objects on the pedestal or only one? It might strike you at first as obvious that there is only one object there. But if this is correct, then this means that Lump = Goliath. And it seems this cannot be right since, using Leibniz's law and the fact that Lump has several properties that Goliath lacks, we can deduce that Lump  $\neq$  Goliath.



Figure 3.2 The Puzzle of the Statue and the Clay



**Modal properties:**

properties having to do with what is possible, impossible, necessary, or contingent

What are these properties? Some are temporal properties. For example, Lump has the property of existing at noon. Goliath lacks this property. Goliath has the property of having been created after 8pm on the day in question. Lump lacks this feature. Lump and Goliath differ as well in their **modal properties**. These are properties having to do with what is possible or necessary, with what could or must happen to these objects if certain conditions were to obtain. While Lump has the property of being able to survive being squashed and rolled into a ball, the statue Goliath lacks this property. If Goliath were squashed and rolled into a ball, it would cease to exist. There would no longer be such a thing as the statue Goliath. In addition, it seems that Goliath has the property of necessarily being a statue, whereas Lump lacks this property. Lump isn't *necessarily* a statue; it could have existed even if it were never shaped into a statue. It could have always just been a lump of clay. And so even though Goliath and Lump share a lot of properties (they are both located in the same place, on the pedestal; they are both shaped like the ancient warrior Goliath; they are both composed of the same atoms), by Leibniz's law we seem forced to say they are numerically distinct. While at midnight, it might look like there is only one object on the pedestal, there are really two objects there: both Lump and Goliath.

Furthermore, if we consider facts about the identity of the statue and the lump, again we seem led to the conclusion that these objects are not identical. For consider the lump at noon. Call this object  $L_1$ . Now consider the lump at midnight. Call this object  $L_2$ . It seems obvious that  $L_1 = L_2$ . Assuming the sculptor used all of the original clay to create the statue, we seem forced to say that the lump of clay at midnight is the very same lump as the one he started with in the morning. After all, simply molding and shaping an object shouldn't make it cease to exist! As we noted earlier, lumps can survive molding. At the same time, it is just as obvious that  $L_1 \neq$  Goliath. After all Goliath is a statue, and at noon there is no statue. So  $L_1$  can't be Goliath. But then if  $L_1 = L_2$  and  $L_1 \neq$  Goliath, then by the fact that identity is an equivalence relation, we must conclude that  $L_2 \neq$  Goliath. We can show this using an argument that employs the method of **reductio ad absurdum** (or reductio for short).

**Reductio ad absurdum:**

the method of proving a claim by showing that the negation of that claim entails a contradiction

*Argument that the Lump at midnight ( $L_2$ ) is not identical to the statue (Goliath)*

1. Assume for reductio that  $L_2 =$  Goliath
2. Lumps of clay do not cease to exist just by being molded
3.  $L_1 = L_2$  (by premise (2))
4. Goliath did not exist at noon
5.  $L_1 \neq$  Goliath (by premise (4))
6.  $L_1 =$  Goliath (by premises (1) and (3) and the transitivity of identity)

7. So,  $L_2 \neq \text{Goliath}$  (by premises (5) and (6) and reductio ad absurdum)

Therefore,

8. The lump at midnight is not the same object as Goliath.

The resulting position is the **Two Object View** defended by the philosopher David Wiggins in his book *Sameness and Substance Renewed*. This is the view that in general, material objects are distinct from the matter from which they are made. We have illustrated the main argument for this view using the example of the Statue and the Clay, but any concrete objects whatsoever that are constructed out of some materials will raise the same issue. We could raise the same issue about a table and the wood of which it is made, your body and its constituting lump of organic material, and so on. The Two Object View is interesting, as it seems to contradict a seemingly plausible principle about material objects: the principle that no two material objects can be wholly located at the same place at the same time.

**Two Object View:** the view that material objects are numerically distinct from the matter of which they are made

### TEXT BOX 3.3

#### Reductio Ad Absurdum

**Reductio ad absurdum** (or simply, **reductio**) is a method of argument often used by philosophers. To apply this method, one assumes the negation of the claim one wishes to prove, and then shows that this assumption leads to a contradiction (both  $P$  and  $\neg P$ ). This, one might say, reduces the negation of the claim to an absurdity, thus showing that one's initial claim is true.

One way of avoiding the Two Object View is to give up on the claim that  $L_1 \neq \text{Goliath}$ , thus challenging premise (4) of the argument above. We might concede that Goliath is a statue of a mythical warrior, and at noon, the lump doesn't look like a statue of a mythical warrior, but really it is. Like the Two Object View, this view has some strange consequences. For it certainly seems that at noon the sculptor could have had a change of heart or become thirsty, and instead used the clay to create a coffee mug. So, Lump could have been used to create a coffee mug instead of a statue. But then if  $\text{Lump} = \text{Goliath}$ , they are numerically the same object, then we would be forced to say that the statue Goliath could have been a coffee mug. This seems wrong. The lump could have failed to be a statue, but could the statue have failed to be a statue? We return to consider other solutions to this puzzle at the end of the present chapter and in Chapter 9 on persistence.

**EXERCISE 3.3****The Statue and the Clay**

What do you think is the most plausible response to the Puzzle of the Statue and the Clay?

**3.3 The Problem of the Many**

Both of the preceding philosophical puzzles about material objects arise when we consider changes that might occur over a period of time: a ship deteriorating over many years and slowly having its parts replaced, a lump of clay being shaped over the course of a day into a statue of a mythical warrior. These puzzles forced us to confront issues about **diachronic identity**, that is, identity over time. This next puzzle we will discuss isn't a puzzle about diachronic identity, but rather concerns what is the case at a single time. In other words, it is a puzzle about **synchronic identity**. It concerns the relation between ordinary material objects and the collections of particles that might be said to constitute them. For this reason, it is called the **Problem of the Many**. The Problem of the Many was introduced by Peter Unger in an article with that name from 1980.

The problem is one that arises for any familiar material object we may consider, but let's focus on the way it arises for the case of human bodies. Take a moment to examine your own body. At first, it may seem to you that there are clear facts about what is a part of your body and what is not. Take a look at a hand. It is easy with your eyes to trace the outlines of this hand and in this way mark the boundaries between what is a part of your body and what is a part of the surrounding environment. But now imagine zooming in closer on your body with a microscope. If you did this, things would appear differently. You could see the cells making up your skin. Zoom in further with a scanning tunneling microscope and you might observe the individual atoms and molecules making up these cells. At this level of magnification, the boundary between your body and the surrounding environment would become much less clear. Maybe some technology one day could allow us to see down to the level of electrons, protons, and neutrons making up these atoms, perhaps even to the quarks making up these protons and neutrons. But we don't need to worry about that. Let's stay at the level of atoms. There are a finite number of atoms that make up your body. Let's suppose this number is (what is supposed to be average)  $7 \times 10^{27}$ .<sup>8</sup> Call the collection of all of these  $7 \times 10^{27}$  atoms 'C.' C is structured in such a way as to make up a human body, you. What Unger points out is that there are also many other collections of atoms just slightly different from C that also seem capable of constituting a human body

**Diachronic identity:**  
identity over time

**Synchronic identity:**  
identity at a time

**Problem of the Many:**  
a philosophical problem about the existence and identity of material objects that arises from the fact that they seem not to have well-defined boundaries; if there are several precisely defined objects with determinate boundaries in the vicinity of a material object, this raises the question of which if any that object is identical to

like yours. And these other collections are located in almost the same place as C. For example, consider a collection of  $7 \times 10^{27} - 1$  atoms, the collection of atoms that includes all of the members of C except for one particular carbon atom at the tip of your little finger. Call this collection 'C-.' This collection of atoms also seems to have what it takes to make up a human body. After all, we could imagine a case in which that atom was never a part of you and we wouldn't be forced to conclude in that case that you lacked a body. Similarly, we can imagine a distinct collection of atoms C--. This collection is just like C except that it fails to include a single hydrogen atom on the inside of your ear. Again, it seems that C-- has everything it takes to constitute a human body. Unger's point is that there are very many collections of atoms located very close to where your body is located and each of these collections has, it would seem, an equally good claim to constitute a human body. And so we seem forced to admit that even here in the same location at which you are sitting, there are a great many other human bodies that look a great deal like you, but still, because they are composed of slightly different atoms, by Leibniz's law, they are not identical to you.

This problem is easily extended to the case of any complex, material object we are familiar with in our day-to-day lives. Just as we argued there are many human bodies in roughly the same place you are sitting, so we could argue there are many chairs in the location under you, many phones, many tables, and so on. Unger's own solution to this puzzle is to deny that there are any complex material objects in the first place. If any collection of atoms has what it takes to make up a human body or a chair, then too many things would have what it takes, so Unger suggests we should revoke our ontological commitment to familiar material objects like human bodies or tables or chairs. If the one commits you to the many, then maybe we shouldn't have believed in the one in the first place.

But perhaps there is a way to distinguish one of these collections of atoms as special, as the one that *really* makes up your body. One thing you might try to argue is that of all of these many collections of atoms, C, C-, C--, and so on, there is one that is the largest - namely C. There is one that includes all of the atoms contained in any of the other collections. One might then identify your body with C, and say that C-, C--, and so on all fail to be bodies, because it is part of the concept of a body that it is the largest such entity of any such sequence. This would be to take the property of being a body to be what Theodore Sider has called a **maximal property**. Sider suggests just such a move in the following passage:

**Maximal property:** a property F is maximal if large parts of an F are not themselves Fs

Ordinary sortal predicates<sup>9</sup> typically express maximal properties, where a property, F, is maximal, roughly, if large parts of an F are not themselves Fs. A large part of a house - all of the house save a window, say - does not itself count as a house. A large part of a cat -

all of it save the tail, say – does not itself count as a cat. Otherwise in the vicinity of every house there would be a multitude of houses; in the vicinity of every cat there would be a multitude of cats. (Sider 2003, p. 139)

This is a way one might try to evade the Problem of the Many, by an appeal to the maximality of the properties of being a house or being a body.

The trouble with this is that just as we may consider the case of the  $7 \times 10^{27}$  atoms making up your body with an atom here or there subtracted, we may also consider the case of the  $7 \times 10^{27}$  atoms plus one or two that someone might otherwise think are part of the surrounding environment. David Lewis, in his discussion of the Problem of the Many focuses on these other potential objects, those that would result from adding atoms rather than subtracting atoms from some initial collection C:

Think of yourself, or any organism, with parts that gradually come loose in metabolism or excretion or perspiration or shedding of dead skin. In each case, a thing has questionable parts, and therefore is subject to the problem of the many. (Lewis 1993, p. 165)

Just as we can consider C- and C-, we might consider also C+, and C++. C+ is C with the addition of an atom in a bead of perspiration that is just making its way from the skin. C++ is C with the addition of an atom that was part of some food that has just begun to be digested by your body. Unger would want to say too in this case that C+ and C++ have just as much claim to constitute a human body as does C. If the subtraction of an atom here or there is insignificant, so is a small addition. But if so, then an appeal to maximality doesn't seem to be able to solve the problem. For which is the maximal entity in this case? Is it C? Is it C+ or C++? C is a part of both C+ and C++. But which is the object that makes the person? There seems no way to tell.

### EXERCISE 3.4

#### The Problem of the Many

There seem to be three ways to respond to the Problem of the Many:

- A. Deny that any of the collections of atoms make up a human body.
- B. Find some way (like Sider) to distinguish one of the collections as special and the only one that constitutes a human body.
- C. Accept the consequence that what seems to be only one human body is really many bodies.

Which of these options strikes you as most promising? Is there any other option available?

### 3.4 The Special Composition Question

So far, in our discussion of all of these puzzles, we have made an assumption. This is an assumption that we now need to call into question, as it has been the subject of much debate among metaphysicians in recent years. What we have been assuming all along is that individual particles (objects of the kind fundamental physics discovers like electrons, quarks, and so on) combine at least some of the time to form collections that are atoms, lumps of clay, hunks of organic material. We have asked about particular collections of atoms, whether they constitute ships or people, but some philosophers think we need an argument that these collections of atoms themselves exist. As Peter van Inwagen notes in his 1990 book *Material Beings*, perhaps individual particles never combine to form composite objects, or perhaps they do sometimes but not always. Maybe, although some particles exist, there is no such thing as the collection of these particles.

In *Material Beings*, the question van Inwagen wants to address is: In what circumstances do some objects compose something? He calls this 'The Special Composition Question.'<sup>10</sup> It is worth emphasizing how it is that our presentation of all of the earlier puzzles and paradoxes made this crucial assumption - that composition occurs at least some of the time. This is most obvious in the case of the first two puzzles (the Ship of Theseus and the Statue and the Clay). There we started by assuming in the first place that there were some planks of wood and that these compose a ship, or we assumed that there was a lump of clay and that at midnight this was sculpted to compose a statue. But even in our presentation of the Problem of the Many, we assumed that composition occurs at least some of the time. After all, we assumed that there were all of these collections of atoms, C, C-, C--, and so on. But to assume there is a *collection* of atoms is to say that the atoms have composed something, if only the collection. And what van Inwagen wants to make us see is that perhaps there isn't any such thing as the collection of atoms, C, C-, or C--. Perhaps just the atoms exist, and these individual things never come together to form an object that is made of them, what we will refer to as their **mereological sum** or their **fusion**.

In order to state the Special Composition Question precisely and accurately, van Inwagen notes that it is necessary to introduce a new kind of logical apparatus, that of plural quantification (1990, pp. 22-26). Following van Inwagen then, we will use ' $\exists x$ s' to symbolize the English phrase 'there exist some xs.' We will also use plural variables 'the xs' to refer to these xs. This will allow us to speak of them plurally without automatically assuming that there is some thing ( $\exists x$ ) that is the sum or collection of the xs.<sup>11</sup> This would assume already that composition does occur.

**Mereological sum (or fusion):** the mereological sum (or fusion) of some objects  $x_1, x_2, \dots, x_n$  is the object that contains  $x_1, x_2, \dots, x_n$  as parts

**Special Composition**

**Question:** the question for any *xs*, when is it the case that there is a *y* such that the *xs* compose *y*

The official formulation of the **Special Composition Question** will be:

For any *xs*, when is it true that  $\exists y$  (the *xs* compose *y*)?

We may state this in quasi-English as: For any *xs*, when is it true that there exists a *y* such that the *xs* compose *y*? Possible answers to this question will all be of the form:

$\forall xs \exists y$  (the *xs* compose *y* iff the *xs* ...)

where the ellipsis (...) is filled in by the condition required to get any *xs* to compose something.<sup>12</sup>

Note how general an issue this is. Answers to the Special Composition Question will tell us what it takes for any objects whatsoever to compose something. Some metaphysicians believe there is a particularly interesting class of objects, **simples** or **mereological atoms**. Simples are objects that have no **proper parts**, where some *x* is a proper part of another object *y* just in case *x* is a part of *y* and *x* is not identical to *y*. (Strictly speaking, every object is a part of itself.<sup>13</sup>) An answer to the Special Composition Question will tell us what it takes for any things whatsoever to compose something, whether those things are mereological simples or mereologically complex objects (objects that are composed by objects other than themselves).

It may help to note another way of understanding the Special Composition Question. Although van Inwagen insists this is not the official formulation of the question, he notes that this more "practical" formulation may help to see the issue at stake:

Suppose one had certain (nonoverlapping)<sup>14</sup> objects, the *xs*, at one's disposal. What would one have to do – what *could* one do – to get the *xs* to compose something? (van Inwagen 1990, p. 31)

Answers to the Special Composition fall under two types: moderate answers and extreme answers. Moderate answers are those that entail that at least some of the time some objects (*xs*) come together to compose a new thing, but also entail that this does not happen all of the time. Extreme answers to the Special Composition are those that entail either that composition occurs all of the time, for any *xs* we consider, or that composition occurs none of the time (at least no distinct objects ever combine to form a new object; everything trivially composes itself). In the remaining sections of this chapter, we will introduce the various answers that have been proposed so that the reader may assess for themselves what to think about this issue van Inwagen raises.

### 3.5 Moderate Answers to the Special Composition Question

There are many different answers that are available to one who would like to answer the Special Composition Question in a moderate way – by giving

**Mereological atom (or simple):** an object lacking proper parts

**Proper part:** *x* is a proper part of *y* just in case *x* is a part of *y* and *x* is not identical to *y*

an answer that entails that composition occurs only some of the time. Moderate answers are attractive to those who would like a metaphysics of material objects that lines up as much as possible with our pretheoretical beliefs about when composition does and does not occur. For example, consider all of the objects of various sizes and complexity that exist here on Earth. Ordinarily, before we go too far in thinking about metaphysics, we think that some of these objects combine to compose more complex objects, while others do not. For example, the bricks that are piled neatly on top of one another at 10 Downing Street in London, England, do compose something: the official home and office of the UK Prime Minister. On the other hand, the fleas on the fattest dog in Scotland and the Statue of Liberty do not combine together to compose anything. There is no such object that is composed exactly of the fleas on the fattest dog in Scotland and the Statue of Liberty. If we really wanted to, we could invent a name, 'Fleabert,' and insist that this name will be used to denote the object composed of the fleas on the fattest dog in Scotland and the Statue of Liberty. But this would just be a way of playing a game with language. Which objects exist is an objective matter and if there is no such object as "Fleabert," then inventing this name will not change that fact. So, one common motivation for seeking out a moderate answer to the Special Composition Question is to find an answer that fits well with our pretheoretical intuitions about which objects exist. The home at 10 Downing Street does exist. Fleabert does not.

One natural attempt at answering the Special Composition Question with a moderate response is provided by the Contact answer. This says:

*Contact:*  $\forall x s \exists y$  (the  $x s$  compose  $y$  iff the  $x s$  are in contact).

The rough idea is: if you have some things, to get them to compose some further object, all one needs to do is bring them next to each other so that they touch.

What is appealing about this answer is that it seems to do well at getting the cases right that are typically used to motivate moderate responses. According to the Contact view, the house at 10 Downing Street exists but Fleabert does not because while the bricks at 10 Downing Street are touching each other, the Statue of Liberty is not in contact with any of the fleas on the fattest dog in Scotland. However, even though there are some cases where the Contact view succeeds, there are other clear cases where it does not.

First, we often think that composition occurs in some cases even though the composing objects are not in contact. For example, the planets Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune together compose our Solar System and yet they are not in contact with one another. Indeed, they are separated by great distances. Van Inwagen also notes that if we should accept any ordinary cases of composition involving medium-sized objects, it seems it first needs to be the case that the fundamental particles of physics come together to compose relatively



basic objects like atoms and molecules. But there isn't good reason to believe that fundamental particles like electrons and quarks compose larger objects like nuclei, atoms, and molecules by coming into contact with one another (1990, p. 34).

In addition, van Inwagen notes that there are clear cases where objects are brought into contact and yet we don't think composition occurs. For example, what happens if you and I shake hands (1990, p. 35)? Now our bodies are in contact and yet there is no reason to think that a new object has come into existence that persists from the moment our hands meet to the moment we let go. But this is a conclusion that is entailed by the Contact view. Thus, it seems we must find some other way of answering the Special Composition Question.

In his book, van Inwagen considers several other moderate answers to the Special Composition Question that he immediately rejects. These are:

*Fastening*:  $\forall x s \exists y$  (the  $x s$  compose  $y$  iff the  $x s$  are fastened to one another, where the  $x s$  are fastened when among the many sequences in which forces of arbitrary directions and magnitudes might be applied, at most a few of them would be capable of separating the  $x s$  without breaking or permanently deforming them).

*Cohesion*:  $\forall x s \exists y$  (the  $x s$  compose  $y$  iff the  $x s$  cohere, where the  $x s$  cohere when they cannot be pulled apart or moved in relation to each other without breaking).

*Fusion*:  $\forall x s \exists y$  (the  $x s$  compose  $y$  iff the  $x s$  are fused, where the  $x s$  are fused when they are joined together such that there is no boundary between them).

All of these improve on the Contact view in some ways, in making the relationship that must obtain between the  $x s$  in order for them to compose something more stable. However, each suffers from new counterexamples.

### EXERCISE 3.5

#### Moderate Answers to the Special Composition Question

For each of Fastening, Cohesion, and Fusion, give one example of a case of composition that would motivate the answer (a case in which composition occurs and the relevant condition is satisfied) and one case that would serve as a counterexample to this answer. The counterexample may either be (i) a case in which the condition is satisfied but composition does not occur, or (ii) a case in which composition does occur but the condition is not satisfied.

Given the difficulty of stating a satisfactory moderate answer to the Special Composition Question, at least one metaphysician, Ned Markosian, has proposed that we instead seriously consider the possibility that there is no good answer to the Special Composition Question. Markosian takes it to be obvious that composition only occurs in some cases and not always, but has argued that van Inwagen's discussion in *Material Beings* supports the claim that one can't formulate a true and interesting answer to this question. The conclusion is that when composition occurs, this is just a brute fact. There is no metaphysical account one can point to that would distinguish the cases in which composition occurs from the cases in which it does not. The house at 10 Downing Street exists and Fleabert does not, but there is nothing more to be said about these bricks or those fleas and that statue that could explain why this is so. And so the only way one could give a true answer to van Inwagen's question would be either to say (trivially) that composition occurs when it occurs, or to list one after the other all of the cases where composition does as a matter of fact occur. Markosian calls this position **Brutal Composition**. More precisely, it is the claim that there is no true, interesting, and finite answer to the Special Composition Question.<sup>15</sup>

**Brutal Composition:** the view that there is no true, interesting, and finite answer to the Special Composition Question

Before we can properly evaluate whether Brutal Composition is right, we should try to see what other answers to the Special Composition Question have been proposed. Then, if at the end of all of this, we see reason to think any answer we attempt to give will fail, we might (reluctantly) come to accept that the facts about composition are brute.

According to van Inwagen, there is something in virtue of which certain *xs* come together to compose a new object, but this isn't to be understood in terms of the *xs*' spatial positions relative to one another. All of the accounts that make whether composition occurs rest on objects' relative spatial positions fall prey to counterexamples. Nonetheless, van Inwagen is certain that at least one mereologically complex material object exists, himself, and so there must be something about the objects that compose him that makes composition occur.

Van Inwagen believes that what makes some mereological atoms compose him is that they participate in a kind of complex activity, the sort of complex activity that allows them to constitute a life (1990, p. 82). So, van Inwagen's proposed answer to the Special Composition Question is:

*PvI*:  $\forall xs \exists y$  (the *xs* compose *y* iff the activity of the *xs* constitutes a life).

What it is for the activity of some *xs* to constitute a life is something that van Inwagen takes to be an empirical matter, something that will be settled by biology and what biologists say is required for some objects to constitute a life. Van Inwagen takes lives to only include those of concrete biological organisms. These arise through the unimaginably complex

self-maintaining behavior of a group of constituents. These groups persist over time by taking in new constituents by ingestion or respiration, and by expelling old constituents.

At this point, I am sure the reader will notice that this view entails quite a surprising ontology. It follows from van Inwagen's answer to the Special Composition Question and the assumption that there are some mereological simples that the only material objects that exist are mereological simples and living creatures. Human beings exist, electrons exist, but tables and chairs, planets and solar systems do not, because they are neither simples nor lives. Certainly, this result is surprising, perhaps too it is counterintuitive if one places weight on the examples usually used to motivate the search for a moderate answer to the Special Composition Question. However, in philosophy just like in science, we often find that the evidence points us in the direction of what is initially counterintuitive. That a view does not agree with our initial intuitions is not a decisive reason against it, but must rather be weighed with all of the other reasons in favor of and against it.

Before moving on to discuss the two extreme answers to the Special Composition Question, it is worth remarking that van Inwagen has gone some way towards showing that even though his view may look surprising at first, it does not necessarily need to conflict with the majority of statements we make in our everyday lives. For van Inwagen notes that there is a very simple way to paraphrase statements that might look to commit one to the existence of mereologically complex objects that are not living (such as tables and chairs) into statements that do not so commit one (1990, pp. 108-111).

For example, consider the sentence 'There are at least two chairs in this room.' At first glance, this seems to commit us to the existence of chairs, and since chairs are mereologically complex material objects that are not living, this seems to commit us to cases of composition violating van Inwagen's answer to the Special Composition Question. A natural first-order symbolization is:

$$\begin{aligned} &\exists x \exists y (((x \text{ is a chair} \wedge x \text{ is in this room}) \wedge (y \text{ is a chair} \wedge y \text{ is in this room})) \\ &\wedge x \neq y) \end{aligned}$$

But there is a paraphrase available. This paraphrase allows us to see the original sentence as only committing us to simples (entities in which van Inwagen does believe) arranged in certain ways:

$$\begin{aligned} &\exists x s \exists y s (((\text{the } x s \text{ are arranged chairwise} \wedge \text{the } x s \text{ are in this room}) \wedge \\ &(\text{the } y s \text{ are arranged chairwise} \wedge \text{the } y s \text{ are in this room})) \wedge \\ &\text{the } x s \neq \text{the } y s) \end{aligned}$$

This regimentation fails to commit one to the existence of chairs. It only commits one to things arranged chairwise. It is not too hard to see how

one might extend this analysis to all other sentences that appear at first to quantify over complex yet inanimate material objects. Van Inwagen may replace talk of tables with talk of simples arranged tablewise, talk of planets with talk of simples arranged planetwise, and so on. In this way, van Inwagen need not say that any of our ordinary beliefs or statements are false. They appear to commit one to complex inanimate material objects, but when understood properly, they do not.

### EXERCISE 3.6

#### Van Inwagen's Proposed Answer and the Method of Paraphrase

How would van Inwagen propose we regiment the following sentences into the language of first-order logic so that they are consistent with his answer to the Special Composition Question?

- A. There are planets made of iron.
- B. Some tables are heavier than some chairs.
- C. Some tables are heavier than some people.
- D. It is not the case that some tables are heavier than some planets.

## 3.6 Mereological Nihilism

We now turn to consider the last two answers to the Special Composition Question. These are the two extreme answers, and it is fair to say these are the most common responses one finds defended in the philosophical literature. Let's begin with **mereological nihilism** (hereafter, nihilism). If one is a nihilist, one answers the Special Composition Question in this way:

*Mereological Nihilism:*  $\forall x s \exists y$  (the  $x s$  compose  $y$  iff the  $x s$  are exactly one).

**Mereological nihilism:** the view that there are no mereologically complex objects, only simples

Strictly speaking, the nihilist doesn't say composition never occurs. The nihilist will allow that any simple object is such as to compose itself. But when the  $x s$  are two or more, there is never anything that the  $x s$  compose. In other words, everything that exists is a mereological atom, a simple.

Extreme though it is, there are reasons to find nihilism compelling. The most obvious point in favor of the position stems from Ockham's Razor. If we are confident that a fundamental scientific theory appealing only to mereological atoms can provide a complete explanation of everything that happens at our world, Ockham's Razor directs one not to posit in addition to these simples mereologically complex objects composed out of them.

Despite this potential benefit in ontological parsimony, there are two arguments one typically finds against nihilism. The first is the one that van Inwagen himself uses to reject the view and it has already been alluded to above. Van Inwagen presents the following argument (1990, p. 73):

*Van Inwagen's Argument against Mereological Nihilism*

1. I exist.
2. I am not a mereological atom.

Therefore,

3. At least one object exists that is not a mereological atom.

Therefore,

4. Nihilism is false.

Nihilists, if they are to maintain their position, must find some way of rejecting this argument. Since the validity of the argument is not questionable, this means finding a way to reject at least one of either premises (1) or (2). When it comes to premise (2), it is difficult to accept that van Inwagen is a material entity of some kind and yet not complex. If van Inwagen is to be identified with any kind of material object, he surely isn't a simple object like an electron or quark, but something that has parts and exists due to the complex behavior of these parts. One possibility is to adopt some form of **mind-body dualism** according to which you and I and van Inwagen are not material objects at all, but rather immaterial minds or souls. This is one way to reject premise (2). However, it is a way that won't be appealing to the naturalist or physicalist philosophers who believe that minds are themselves parts of the material world, arising from complex processes of physical matter. Another option for the nihilist is to deny (1) and say the same thing about van Inwagen as he himself says about tables and chairs. The nihilist can then say while this claim is false:

Van Inwagen exists,

this claim is true:

There exist some simples arranged van-Inwagen-wise.

That is, for all claims van Inwagen accepts that look to commit him to his own existence, the nihilist can produce paraphrases that show them only to be committed to simples. This itself is controversial. However, debate on this topic continues.<sup>16</sup>

The second argument that many have found to give a good reason to deny nihilism has to do with the fact that for all we know, there might be no bottom level of reality. After all, how can we be sure right now that there really exist some objects that are, as a matter of fact, mereological

**Mind-body dualism:**  
the view that there are two kinds of substances, minds (mental substances) and bodies (material substances)

atoms? Given today's physics, we may think that electrons and quarks are mereological simples; they don't have any proper parts. But time and again, throughout the history of physics, when physicists thought they had reached some realm of ultimate mereological atoms, it turned out that there was more structure to discover. Indeed, traditionally the word 'atom' refers to what is indivisible (*a-tomos*). In the early nineteenth century, John Dalton introduced the term to refer to what he then thought were the mereological atoms, and the term stuck. We now know however that these objects, what we today call 'atoms,' are not really simple at all. And so, what reason is there to think the basic entities of today's physics are the real mereological atoms?

The problem this causes for nihilism is that if there are no mereological simples, but everything is ultimately mereologically complex all of the way down, then it follows from the nihilist's view that no material objects whatsoever exist. After all, the nihilist's view is that only simples exist. This seems problematic. We might question whether there are ultimately only electrons and quarks or whether there are tables and people too, but surely there are material objects of some kind. If nihilism is right, then maybe this isn't so.

We should note that a similar problem obtains for van Inwagen's view as well.<sup>17</sup> If it turns out that there are no simples and van Inwagen's view is correct, then the only things that will exist are living things. There will be no electrons, quarks, tables, or chairs. The only things that will exist are those things that are alive. Since we know this scenario to be false and yet we don't know whether matter is infinitely divisible, it appears we must conclude that van Inwagen's view is false as well.

### 3.7 Mereological Universalism

The last answer to the Special Composition Question we will consider is what is arguably the most common response. This is:

*Mereological Universalism:*  $\forall x_s \exists y$  (the  $x_s$  compose  $y$  iff the  $x_s$  are disjoint).

To say the  $x_s$  are disjoint is to say that they do not spatially overlap (i.e., their spatial locations are entirely distinct). This is a view that has been given an influential defense by David Lewis in his book *On the Plurality of Worlds* (1986). Lewis states the position this way:

I claim that mereological composition is unrestricted: any old class of things has a mereological sum. Whenever there are some things, no matter how disparate and unrelated, there is something composed of just those things. (Lewis 1986, p. 211)

**Mereological**

**universalism:** the view that any material objects, so long as they do not overlap spatially, compose some further object

This view entails that for any material objects whatsoever (simple or complex), as long as they do not overlap spatially, there is something that these objects compose. Some philosophers, for example, Markosian, argue that this view should be rejected for coming into too much conflict with our background beliefs about when composition occurs and when it doesn't. To return to our earlier examples, there is no such thing as Flea-bert; there is no such thing as the object composed of the atoms making up you and me when we shake hands. **Mereological universalists** (universalists, for short) typically dismiss this sort of worry by insisting that we shouldn't rely so heavily on our pretheoretical intuitions about when composition occurs and when it doesn't. We shouldn't use our intuitions to guide our reasoning about what sorts of things do and do not exist when we have no reason to believe our intuitions will lead us toward the truth. Would following our intuitions have ever led us to quantum mechanics or general relativity, the most well-confirmed theories we have ever had?

But it is worth pointing out that even if the appeal to intuitions is not in general a reliable method of coming to the truth about what exists, in this particular case, there is something to be said for respecting them. The reason is that what universalists are asking us to believe is that composition occurs no matter which group of material objects we consider. And yet the way we come to understand what it is for some things to compose some other thing in the first place has always involved us drawing distinctions between the cases in which it occurs and the cases in which it does not. But what could it mean to say that every case in which some objects exist is a case of composition? Could we ever even understand such a claim? Doesn't this require giving up on our usual understanding of what 'composition' means? And if so, is this really an answer to the Special Composition Question or just a change in topic?

In addition to these worries about the conflict of the view with our intuitions, there is of course the fact that the position involves a considerable cost in ontological parsimony. The universalist's ontology is enormous and certainly larger than that of any view about material objects we have considered up until now.<sup>18</sup> Still, there is at least one argument that motivates many metaphysicians to endorse it.

This argument relies on the premise that if one doesn't say that composition always occurs (or never occurs, but for now let's set aside nihilism), then one is forced to admit the existence of cases in which whether composition occurs or not is vague. But if it is vague in a given case whether composition occurs, then it is vague what exists. After all, when we are considering a case of composition involving multiple *x*s, if the *x*s compose *y*, then there will be an additional object *y* that is the sum or fusion of

the *xs*. And so, if it is vague whether the *xs* compose *y*, then it is vague whether *y* exists. But existence isn't vague. There is no way to just kind of exist. Either something exists or it doesn't.<sup>19</sup>

Why would someone with a moderate view (i.e., a view that rejects both universalism and nihilism) have to admit that there are cases where it is vague whether or not composition occurs? The reason is that every moderate answer that has been proposed to the Special Composition Question uses vague concepts. Contact, fastening, cohesion, fusion, involved in a life: all of these are vague notions. And since they are vague, we can conceive of cases where it is indeterminate whether or not an object satisfies them. Let's examine this for the concept of fastening. Recall, by definition, objects are fastened when among the many sequences in which forces of arbitrary directions and magnitudes might be applied, at most a few of them would be capable of separating the *xs* without breaking or permanently deforming them. But who is to say what counts as "a few of them"? This is why the notion of fastening is vague. All of the other notions are vague as well. Indeed, van Inwagen concedes that the concept of a life is vague. There are boundary cases in which there is no objective fact about whether some *xs* constitute a life or not. To see this, one only need consider the earliest stages of an organism after conception. At what point does the life begin?

Indeed, one of the main reasons why metaphysicians have been skeptical about moderate answers to the Special Composition Question is that these answers all involve vagueness. Of the responses to the Special Composition Question, only the extreme answers (nihilism and universalism) and the Brutal Composition view eliminate vagueness about when composition occurs (and so how many objects exist) by saying that composition always occurs, it never occurs, or it is simply a brute fact when and where it occurs. Since this issue of vagueness is so central to debates about composition and indeed many others in metaphysics, we will devote the next section of this chapter to the topic.

### EXERCISE 3.7

#### Responding to the Special Composition Question

We have now seen several moderate and two extreme answers to the Special Composition Question. What is your preferred answer? How would you defend this response against the worries raised earlier?



### 3.8 Vagueness

It was Lewis who originally appealed to vagueness to raise a problem for moderate answers to the Special Composition Question. We should note that he was not denying that there is such a phenomenon of vagueness in our world. Everyone grants that there are many concepts that form part of our ways of thinking about the world and words that form parts of our language that are vague. What Lewis wanted to insist, however, is that vagueness is *merely* a phenomenon of our language or our way of conceptualizing the world. There is no vagueness in the world as it is in itself, independently of our ways of thinking or talking about it. While Lewis had no problem with what is typically called **linguistic vagueness**, what he was objecting to was **metaphysical vagueness**:

**Linguistic vagueness:** vagueness that is the result of semantic indecision; there not being facts to determine precisely in all cases what our terms apply to

**Metaphysical (or ontic) vagueness:** vagueness that results from how the world is objectively, not how we think or talk about it

The only intelligible account of vagueness locates it in our thought and language. The reason it's vague where the outback begins is not that there's this thing, the outback, with imprecise borders; rather there are many things, with different borders, and nobody's been fool enough to try to enforce a choice of one of them as the official referent of the word 'outback.' Vagueness is semantic indecision. (Lewis 1986, p. 212)

According to many (though not all) philosophers, metaphysical vagueness does not just fail to exist at our world, it is unintelligible. What could it mean to say that whether or not a certain object exists is vague? That it only partially or somewhat exists? If an object exists in any way or to any degree, then doesn't it exist?

Another option for how to understand vagueness that is compatible with these worries about metaphysical vagueness, but doesn't understand vagueness as the result of semantic indecision, is a view called **epistemicism**. According to epistemicism, it is determinate what our words mean. Our linguistic behavior and dispositions fix what we mean by 'life,' 'the outback,' and so on. It is just that these facts about meanings are not completely known to us. Applying epistemicism to the case of composition, one would say we simply can't know in every case whether composition occurs. This doesn't mean there isn't a fact about when composition occurs and when it doesn't. It is just that these facts are sometimes hidden from us. We can tell in the easy cases that composition occurs or it doesn't. But in those borderline cases that Lewis is worried about, we are unable to tell.<sup>20</sup>

Whether vagueness is ultimately to be understood as semantic indecision or as a failure on our part to know the right thing to say about borderline cases, many believe it is ultimately a phenomenon having to do with our (human) connection to the world. It is not a feature of what the world is like independently of us. And so, if one has trouble understanding how there can be vagueness in the world, then one should either try to improve

**Epistemicism:** the view that vagueness is a form of ignorance; it is not a matter of indeterminacy in the world or in what our words or concepts apply to, but our ignorance about what our words or concepts apply to

on the formulation of moderate responses so as to eliminate reference to vague terms, or one should prefer some other response to the Special Composition Question.

On the other hand, one could also do more work to try to make sense of how there could be vagueness in the world, in order to bolster the case for the moderate responses. Metaphysicians like Ken Akiba, Elizabeth Barnes, J. Robert G. Williams, and Jessica Wilson have each developed accounts of what it could mean for there to be genuine metaphysical or ontic vagueness, rather than just vagueness in our thought and language, what it would mean, to return to Lewis's case, for the outback itself to be vague.

Here, for example, is Akiba's proposal for how to understand metaphysical vagueness:

Many people think that the material world consists not only of three dimensions of space, but also of the fourth dimension, i.e., the temporal dimension. . . . According to the view I advocate, the entire world has another dimension: the *precisificational* dimension. This dimension, which concerns determinacy and indeterminacy, is made up of *precisified* worlds, in which everything is precise . . . every material object extends over precisified worlds, and two objects may coincide with each other in one precisified world but diverge in another. (Akiba 2004, p. 409)

Akiba is here positing a multiverse of what he calls "precisificational worlds." In each precisificational world, all facts are determinate. However, since objects are spread over multiple precisificational worlds, and they might have different precise characteristics at different such worlds, these objects can themselves be vague. According to this account, the outback is a vague object just in case it occupies one precise region of space *R* in one precisificational world; but a distinct precise region, *R'*, in another precisificational world. The outback fails to be a vague object just in case it occupies the same precise region of space in all precisificational worlds.

Note that the way Akiba presents this view, all of the precisificational worlds are equally real. Just as we believe that entities at other spatial locations are just as real as entities over here, entities at the different precisificational locations in Akiba's model are equally real. On this view, what it would mean for composition to be vague would be for some *xs* to compose an object *y* in some precisificational worlds but not others.

Akiba's proposal is fascinating and of course immediately raises the question of how plausible it is to believe that there is this additional dimension to reality. Akiba concedes that physicalists may be skeptical toward the proposal - is this how we may discover extra dimensions to reality? By a priori argument? But he argues that rival approaches also have ontological costs. Overall, where things stand today, what Akiba and others

who propose accounts of metaphysical vagueness have done is give us some conceptual path to understanding what it could even mean to say that there is vagueness in the world, not just in our thought or language. And this is an important first step in overcoming the challenge vagueness raises for moderate approaches like van Inwagen's.

### 3.9 Back to the Paradoxes

We have now considered several different answers to the Special Composition Question. Although we have seen the usual lines of arguments that get brought out in support of or against the various views, I haven't attempted to convince you that you should adopt one of these views over the others. All of the views have costs. Now that all of the views have been described, we can briefly come back to the paradoxes of material constitution with which we began this chapter and see if the various views shed any light on them.

Interestingly, van Inwagen defends his own response to the Special Composition Question by noting that it allows him an easy way to get out of the paradoxes of material constitution. For those involving the persistence or constitution of inanimate material objects, he can say simply that there is no paradox because none of these objects exist. There is no Ship of Theseus; there are no statues, no lumps. The mereological nihilist as well can give a similar response to these paradoxes and say they give one all the more reason to think there are no such things as mereologically complex material objects.

When it comes to paradoxes that don't involve inanimate objects, but instead living ones – for example, we can ask not about the statue and the lump of clay, but about you and the lump of matter that composes you, and the Problem of the Many was already formulated in terms of living human beings – van Inwagen must give a different response. (Though the nihilist is once again in the clear.) In these cases, van Inwagen thinks that the puzzles arise because one is assuming mereological universalism. For example, the Problem of the Many arises only because we begin by assuming there are all these collections of atoms, *C*, *C*-, *C*++, and so on. Since van Inwagen denies that all of these collections of atoms exist, he can deny there are many human beings in the same approximate place at the same approximate time. There is just the one collection of atoms that comes together to actually constitute a life. For van Inwagen, it is vague what exactly this object is (whether it is *C*, *C*-, or . . .), but this doesn't mean there are many human beings in approximately the same spatial location at the same time. (If we accept Akiba's approach to metaphysical vagueness, we will say rather that there is just one human being, in many different "precisificational" locations.)

When we apply the puzzle of the Statue and the Clay to the case of a human being and its constituent matter, again van Inwagen will be able to argue that his view has an advantage over the mereological universalist's. To see this, consider the intuitions that drove the paradox.<sup>21</sup> We noted that the statue and lump have different temporal and modal properties. While the lump existed at noon, the statue did not. The lump could survive being squashed. The statue could not survive being squashed. By Leibniz's law, following Wiggins, we concluded that the lump and statue must be distinct. To apply these intuitions to the case of a living organism like a person, we might say: while the hunk of matter that composes you existed a hundred years ago (it may not have taken the shape of a human, but it existed), you did not.<sup>22</sup> While this hunk of matter could survive being squashed so that its vital organs stop functioning, you could not survive being squashed so that your vital organs stop functioning. Therefore, applying Leibniz's law, we conclude that there are two objects: you on the one hand, and the hunk of matter that composes you on the other.

Van Inwagen will deny there is any hunk of matter that has these temporal and modal properties. The hunk of matter did not exist before you were born (unless what is unlikely is the case, that the hunk of matter that now makes you up previously made up a different organism). This is because there are no such things, according to van Inwagen, as inanimate hunks of matter. In addition, it is false (at least according to van Inwagen's view) that the hunk of matter that composes you could survive being squashed so that its vital organs stopped functioning. If this happened, then the matter would cease to constitute a life and so it would fail to exist. So, one who wants to defend van Inwagen's view can respond to this paradox of material constitution as well.

When it comes to most of the puzzles of material constitution, perhaps it is really only the other moderate views (besides van Inwagen's) and mereological universalism that seem to face problems. In Chapter 9, we will have reason to return to these paradoxes and see a way mereological universalists typically respond. This will involve a particular view about what it is for objects to persist over time.

## Suggestions for Further Reading

An excellent volume discussing the paradoxes of material constitution and presenting many different solutions to them is *Material Constitution*, edited by Michael Rea. In addition to the works cited above, *Ordinary Objects* by Amie Thomasson and *The Structure of Objects* by Kathrin Koslicki are also recommended. Both present quite interesting views about the nature of ordinary material objects different than those discussed in this chapter. Trenton Merricks's *Objects and Persons* is an excellent defense of a view of material objects in many respects similar to van Inwagen's.

On the issue of whether there could fail to be any mereologically simple objects, but instead complexity all of the way down, Jonathan Schaffer's "Is There a Fundamental Level?" is recommended, as is the discussion in Theodore Sider's 2011 book *Writing the Book of the World*, pp. 133-137. Many of the articles in the volume *Metametaphysics*, edited by David Chalmers, David Manley, and Ryan Wasserman consider the question of whether the debate over the Special Composition Question is a substantive issue.

For more on vagueness, a good collection of essays is *Vagueness: A Reader*, edited by Rosanna Keefe and Peter Smith. Williamson's book *Vagueness* also provides an excellent overview of several rival positions. An influential, but rather technical argument against metaphysical vagueness may be found in Evans (1978). In addition to Akiba's (2004) account discussed in Section 3.8, rival accounts of metaphysical vagueness may be found in Barnes (2010), Barnes and Williams (2011), and Wilson (2013).

## Notes

- 1 See Wilson (2006).
- 2 Some philosophers have expressed concern that the notion of a material object we receive from physics is actually more poorly understood than many metaphysicians realize. For discussion, see Barbara Montero (1999).
- 3 This means they believe there exist at least some concrete material objects and that the existence of these objects does not depend on any human being's thinking about them; they are mind-independent in this sense.
- 4 Please refer back to the discussion of numerical identity or "identity in the strict sense" in Text Box 1.2.
- 5 Mark Balaguer has suggested this argument.
- 6 We will talk more about the meaning of the box ( $\Box$ ) in Chapter 10.
- 7 This also introduces the problem of having to answer the question of *when* it was that the Ship of Theseus ceased to exist.
- 8 In other words, 7,000,000,000,000,000,000,000,000,000.
- 9 A **sortal predicate** is one that classifies an object as a member of a certain sort (or kind). Examples are: 'is a chair,' 'is a planet,' 'is a statue,' and 'is a person.'
- 10 This issue was posed originally in a 1981 paper by H. Scott Hestevold called "Conjoining." There, Hestevold asked: "*When* is it . . . that two objects *do* strictly make up an individual thing?"
- 11 Van Inwagen also raises the same concerns about the terms 'aggregate,' 'array,' 'group,' and 'multiplicity' (p. 22). He rejects the option of using quantification over sets to express the Special Composition Question, but for a different reason. To speak of 'the set of' of some things doesn't assume composition has occurred. Composition is a relation whereby concrete objects combine to create concrete objects, not abstract objects like sets. Nonetheless, van Inwagen notes the assumption of abstract objects like sets isn't necessary to raise the Special Composition Question or formulate an answer to it. We can do this without entering the debate about abstract entities.
- 12 'iff' is an abbreviation for 'if and only if.'
- 13 This is according to the formal theory of mereology that many metaphysicians use to understand the notion of parthood, that of Henry S. Leonard and Nelson Goodman (1940).
- 14 By 'nonoverlapping,' we mean that the objects do not have any parts in common.

- 15 Although Markosian famously defended Brutal Composition in his 1998 paper with that name, he now prefers a different response to the Special Composition, one he calls 'Regionalism.' See Markosian (2014) and Gilmore and Leonard (2020) for discussion.
- 16 For example, see the defense of nihilism in Sider (2013).
- 17 This was noted by Sider in his 1993 paper "Van Inwagen and the Possibility of Gunk." Gunk is a technical term introduced by metaphysicians to refer to a material object that has no mereological simples as parts. Instead, its proper parts are such that each of their proper parts in turn have proper parts.
- 18 Though does it commit us to more *types* of things? Recall from our discussion of Ockham's Razor in Chapter 1, this is the real issue.
- 19 This argument from vagueness against moderate answers to the Special Composition Question and in favor of mereological universalism was originally presented in Lewis (1986, pp. 212-213) and developed further in Sider (2001, Section 4.9).
- 20 Epistemicism was defended by the philosopher Timothy Williamson in his 1994 book *Vagueness*.
- 21 See *Material Beings*, pp. 75-79.
- 22 When we talk about 'this hunk of matter,' note we are talking about the mereological sum of the atoms that compose you. This is not a plural-referring term like 'these atoms.'

# 4 Critiques of Metaphysics

## Learning Points

- Introduces logical positivism
- Presents two critiques of metaphysics from the logical positivist school and replies
- Presents a more recent critique of metaphysics from a naturalist perspective and replies
- Evaluates the relationship between metaphysics and science

## 4.1 A Concern about Methodology

So far, we have introduced some of the central questions of metaphysics: ontological questions about the existence of abstract and material objects. In later chapters, we'll address other ontological issues such as whether we should believe in the existence of past or future objects and events or only present ones, whether we should believe, in addition to natural features and objects, in those that are socially constructed like genders or races. We'll also address questions that are not ontological, but more broadly metaphysical in that they concern other features of our world such as the nature of causality or whether we have free will.

Perhaps now, since we've already spent time on some very abstract issues and we've seen arguments on both sides of each one, you may have started to wonder whether there is any satisfactory way of settling these issues. Could there really be a fact about whether there exist universals or only concrete objects? Could there really be a fact about whether mathematical truths involve the existence of a domain of objects inhabiting some realm outside of space and time, or whether they only require our being capable of proving certain claims and not others? What could settle

such issues after all? It seems that philosophers can come up with arguments favoring both sides of every metaphysical debate and these issues are so abstract that it is difficult to see how anything we could do could settle them. And if there isn't any way to settle them, maybe they aren't about anything of substance after all? The Austrian philosopher Ludwig Wittgenstein (1889-1951) famously suggested that the best way to settle philosophical problems, including these metaphysical ones we've engaged here, was through therapy (1958, p. 255); what is needed is not answers to these questions, but a way to get oneself to stop asking them in the first place.

If you've had these worries yourself, you are not alone. As we will see in this chapter, such a worry about metaphysics was quite prevalent at one time in the early twentieth century. It originated in a philosophical movement called **logical positivism**. This movement originated with a group of philosophers and scientists who met in Vienna in the 1920s and 1930s, the Vienna Circle. We will focus in this chapter especially on the arguments of one prominent member of this circle, the philosopher Rudolf Carnap (1891-1970).

**Logical positivism:** a movement in philosophy originating in Austria and Germany in the 1920s; a movement critical of metaphysics, arguing all knowledge must originate from sense experience or logic

## 4.2 Carnap's Two Critiques of Metaphysics

In his "Elimination of Metaphysics through the Logical Analysis of Language" (1932), Carnap argued that the main problem with metaphysics was the distortions of language that metaphysicians use in order to ask their questions and state their theories. These distortions of language are so severe that in the end, Carnap thought all distinctly metaphysical claims were meaningless.

Carnap diagnoses several ways in which these distortions occur. Sometimes this comes about because the metaphysician introduces new words without ever supplying clear meanings for them. Carnap gives the example of 'essence.'<sup>1</sup> Other times, the metaphysician uses words that normally have a clear meaning in an unorthodox and unclear way. One of Carnap's main examples here is 'God.' This word could, Carnap notes, be used to denote some physical being. In ancient times, the word was used with such a meaning to refer to physical beings living on Mt. Olympus possessing special powers. However, today the word 'God' has what for Carnap is a purely "metaphysical" usage. It is used with the intention of denoting something that is not a physical being but lies beyond what could be observed.

When Carnap claims that words are being used so that the sentences they compose lack meaning, he is very explicit about the theory of meaning he intends. To illustrate this theory of meaning, Carnap says that for



a given sentence *S*, the following are all reformulations of the same question (1932, p. 62):

Under what conditions is *S* supposed to be true, and under what conditions false?

How is *S* to be *verified*?

What is the *meaning* of *S*?

Carnap is here endorsing a verificationist theory of truth and of meaning.

The **verificationist theory of meaning** says that a sentence's meaning is given by its means of verification. The **verificationist theory of truth** says that a true sentence is one that is or may be verified. If a sentence cannot be verified, then it is meaningless and lacks a truth value. According to the logical positivists, there are two kinds of verification: verification by sense experience or empirical observation (**synthetic** means) and verification by logical methods (by **analytic** means), where the latter includes mathematical reasoning. If a claim cannot be proven either using logic (or mathematics) alone or empirical observation, then it is unverifiable and hence meaningless.

**Verificationist theory of meaning:** the meaning of a statement is given by its verification conditions

**Verificationist theory of truth:** a sentence is only capable of truth or falsity if it is able to be verified or falsified

**Synthetic:** empirically verifiable

**Analytic:** truth in virtue of logic or meaning of the terms involved

#### TEXT BOX 4.1

##### Methods of Verification

Logical positivists differed on what verification must involve in any particular case, but a pillar of logical positivism was the view that there are two basic kinds of verification: by analytic and by synthetic methods.

When it comes to verification by analytic methods, Carnap and the logical positivists held that this sort of verification involved logic alone. In some cases, a statement is true in virtue of the basic concepts involved in the statement, e.g.:

All triangles have three sides.

In other cases, a statement's verification requires a deductive proof. Carnap himself was inspired by the developments in logic by Frege in the late nineteenth century to think that all mathematical truths could be verified on the basis of deduction from the principles of logic alone. This view, that mathematics is in principle reducible to logic, is called **logicism**.

When it comes to synthetic verification, many positivists thought an ideal language contained a basic set of statements that may be verified directly by sensory experience. These are what Carnap

**Logicism:** the view that mathematics is reducible to logic

called **protocol statements**. For example, a protocol statement might be:

The bell is ringing.

This can be directly verified by hearing the sound of a bell. Still other statements may be verified indirectly, that is through a chain of verification leading ultimately back to protocol statements. For example:

Someone is at the door.

This may be verified indirectly by combining the statement “The bell is ringing” with other background information that the ringing of the bell entails that someone is at the door. Or, if one had already verified the statement that the mailman comes every day at about this time, then the protocol statement “The bell is ringing” could indirectly verify:

The mailman is at the door.

This is a claim that may not be directly verified by experience. One might think you cannot directly hear that the mailman is at the door, only that a bell is ringing.<sup>a</sup> Still one can verify this statement through a chain ultimately leading back to what can be directly verified by experience (see Figure 4.1).

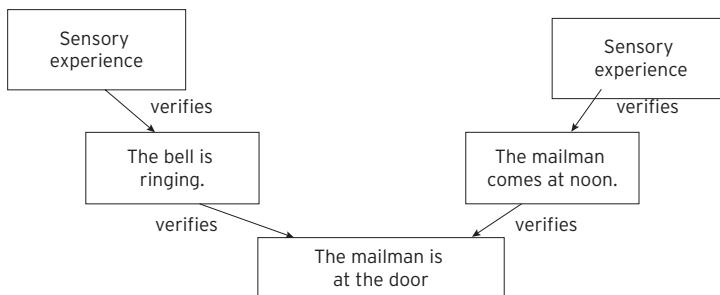


Figure 4.1 A Chain of Verification

In general, the only statements that are meaningful for the verificationist are, on the one hand, the synthetic statements – those that may be directly verified by sensory experience or verified indirectly through a chain of verification leading ultimately back to sensory experience – or, on the other hand, analytic statements, those verified by logic alone.

**Protocol statement:** a statement that may be directly verified by sense experience

So often, the problem with metaphysics is that its statements lack meanings. In a final case Carnap considers, the problem is one of the form or structure of the metaphysical statement (its syntax). Sometimes, Carnap notes, it is not that the metaphysician uses a word that doesn't have a clear verificationist-style meaning, but instead that she is using a perfectly meaningful expression in a perverse way so as to form a sentence that cannot be evaluated for truth or falsity. In one of the German philosopher Martin Heidegger's (1889-1976) most famous essays, "What is Metaphysics?", he offers the view that it is only by contemplating nothingness that the nature of being can really be understood. Trying to say more about this nothing, Heidegger concludes "*Das Nichts selbst nichtet*" or "The nothing nothings." Carnap takes this claim in for an extended bashing. He argues the meaninglessness of Heidegger's thesis consists in its distorting the logical role of the word 'nothing.'

To see this, note what we are saying when we say something using the word 'nothing,' for example, "Nothing is free." Here, we are not using 'nothing' as a noun phrase that refers to some object that can then be some way or have certain properties. This would be to mistake the sentence's logical form for:

Fn.

This symbolization is incorrect. The role of 'nothing' in a sentence is not to serve as a name, but as a quantifier phrase. The logical form of "Nothing is free" is:

$\neg \exists x Fx$

Heidegger's claim "The nothing nothings" thus is ill-formed. It cannot even be expressed in symbolic logic. This is because it treats a quantifier phrase as if it were a referring term. One might try to symbolize it as:

Nn,

if one likes, but this is a distortion of logic. There is no way to symbolize it using correct logical notation. We must conclude then, Carnap says, that Heidegger's sentence is meaningless.

According to Carnap, the trouble with metaphysicians' use of language whether through distortions of logic or meaning, is that such distortions yield sentences incapable of verification. How could one verify if the nothing nothings or a person has an essence or an invisible God exists? Insofar as such questions are unverifiable, they are meaningless.

### EXERCISE 4.1

#### The Verificationist Theory of Meaning

Which of the following claims would count as meaningless according to the verificationist theory of meaning? Explain how they would be verified, or if they are unverifiable, then say so.

- A. There are prime numbers greater than three.
- B. There are numbers.
- C. There are two elevators in the Empire State Building.
- D. The universe began before the Big Bang.
- E. There is intelligent life on other planets.
- F. The God of *Genesis* exists.
- G. Dinosaurs once roamed the Earth.

In “Empiricism, Semantics, and Ontology” (1950), Carnap addresses metaphysical statements that wouldn’t obviously fall prey to the earlier critique: ontological statements like ‘There are material objects,’ ‘There are universals,’ or ‘There are numbers.’ Let’s say that the metaphysician has stripped her language of dubious terms, terms that she cannot supply with clear meanings. And she does her best to make statements that are expressible in clear logical form. Indeed, contemporary metaphysicians do their best to ensure that their statements and theories are always formulated in the clearest terms possible. Nonetheless, Carnap thought there would still be problems for metaphysics.

To see Carnap’s objection, we must start with his notion of a framework. It is always from within a framework that meaningful questions may be asked. It is from within a framework that statements may be evaluated for truth or falsity. When Carnap talks about a **framework**, he has in mind a linguistic framework. Frameworks involve two things:

1. A list of the expressions for the language, and syntactic rules, i.e., rules for how these expressions may be combined to form grammatical sentences

For example, if we are considering the framework of mathematics, say basic arithmetic, the expressions will be numerals like ‘1,’ ‘2,’ and so on, as well as symbols for addition, ‘+,’ ‘=,’ etc. There will be rules that say you may combine these terms in certain ways to yield grammatical sentences, for example ‘ $1+1 = 2$ ,’ and not in other ways such as ‘ $1+=2$ .’

#### Framework (Carnapian):

a linguistic system including rules of grammar and meaning

The second component of a framework for Carnap is:

2. Rules that allow one to evaluate whether or not a given sentence in the language is true or false.

As we have seen, Carnap and the logical positivists endorsed a very specific view about what kinds of rules these could be. This is summarized by the verificationist theory of meaning: a sentence's meaning is given by its rules of verification. Returning to our case of arithmetic, the sentence ' $1+1 = 2$ ' is true just in case the rules of arithmetic allow one to verify this sentence. In an axiomatized mathematics, this is accomplished via the method of proof. And again, there are two broad kinds of verification: verification by analytic methods, which includes logical or mathematical deduction, and verification by synthetic methods, that is, methods that involve empirical observation. In a linguistic framework like that of arithmetic, the semantic rules are entirely analytic. Other frameworks Carnap considers, such as those containing material object terms like 'desk' or 'chair' or 'electron,' will include semantic rules that are synthetic. In this kind of framework, questions will be settled by means of empirical observation.

Given this concept of a framework, Carnap says that all questions we may ask or statements we may make may be viewed as internal or external. An **internal question** is one that is asked and evaluated from within a specific linguistic framework. And an **internal statement** is one that is stated and interpreted from within a linguistic framework. Put this way, we can see that only internal questions and statements are meaningful because only in the context of a framework can we presuppose that there are certain terms and rules for combining them in such a way to form statements capable of being true or false. **External questions** and **external statements** are those interpreted from outside a particular linguistic framework. As such, they are meaningless and not capable of truth and falsity.

Carnap asks whether we should view the questions asked by metaphysicians, in particular, ontological questions, as internal or external questions. Ontological questions are questions like 'Are there numbers?' or 'Do material objects exist?' These questions, as we have seen, can only be meaningful if they are asked as internal questions. The trouble, Carnap argues, is that even if these questions may be evaluated as having answers within a framework with clear syntactic and semantic rules, these ontological questions will always turn out to be hopelessly trivial. As such, they cannot be attributed the significance the metaphysician intends for them. These questions are trivial because once one chooses to use a given linguistic framework, the answers to ontological

**Internal question:** a question asked and evaluated from within a specific linguistic framework

**Internal statement:** one that is stated and interpreted from within a linguistic framework

**External question:** a question attempted to be interpreted from outside a particular linguistic framework

**External statement:** one attempted to be interpreted from outside a particular linguistic framework

questions are always obvious. Carnap uses the following example to illustrate this point.

Let's say one decides to use the framework of arithmetic in order to understand the following two statements and decide whether they are true:

There are prime numbers greater than 1,000,000.

There are numbers.

In order to know whether the first is true, one has to do some arithmetic. One will need to verify the first statement using the analytic methods of the arithmetical framework. One has to actually prove some things. By contrast, one finds that the second sentence is true automatically once one chooses to adopt the arithmetical framework. In choosing the framework, we've chosen to use words like '1' and '2' and it is trivial that these terms will refer to numbers.

To use another example, suppose one chooses to adopt a language that uses terms for material objects. Consider the following two statements:

There are two elevators in the Empire State Building.

There are material objects.

Again, once we presuppose the linguistic framework of material objects, once we, in Carnap's terminology, decide to interpret these statements *internally*, we find that although the first statement is substantive, the second is trivial. To evaluate the truth of the first, we have to apply (in this case) synthetic methods. We will have to engage in some empirical observation, trek out to Manhattan or at least look things up online in order to see whether this claim is true. By contrast, once we have decided to adopt a linguistic framework containing terms for material objects, the second statement, and this is the distinctively ontological one, one that would only be questioned by a philosopher, is trivially true.

So, ontological statements turn out all to be trivial when considered as internal statements. Is there any way to make sense of them so that they are not trivial? We have already noted that if one views them as external statements, they will be meaningless. Yet Carnap concedes there is a good question one can ask in the ballpark of such questions as 'Are there numbers?' or 'Are there material objects?'; namely, whether or not it would be wise to adopt a linguistic system in which such claims come out as true. This is a question that is about the linguistic system as a whole. Although these questions do not have answers that may be evaluated as true or false, since such verdicts can only be reached from within a linguistic system, Carnap does believe that they have answers that may be evaluated in practical terms.

Carnap ultimately summarizes his attitude towards ontological statements, interpreted externally, in the following way:

An alleged statement of the reality of a system of entities is a pseudo-statement without cognitive content. To be sure, we have to face at this point an important question, but it is a practical, not a theoretical question; it is a question of whether or not to accept the new linguistic forms. The acceptance cannot be judged as being either true or false because it is not an assertion. It can only be judged as being more or less expedient, fruitful, conducive to the aim for which the language is intended. (Carnap 1950, p. 214)

This isn't too distant a verdict from that reached in the earlier 1932 paper:

The (pseudo)statements of metaphysics do not serve for the *description of states of affairs*, neither existing ones . . . nor non-existing ones. . . . They serve for the *expression of the general attitude of a person towards life*. . . . What is here essential for our considerations is only the fact that art is an adequate, metaphysics an inadequate means for the expression of the basic attitude. Of course, there need be no intrinsic objection to one's using any means of expression one likes. But in the case of metaphysics we find this situation: through the form of its works it pretends to be something that it is not. The form in question is that of a system of statements which are apparently related as premises and conclusions, that is, the form of a theory. In this way the fiction of theoretical content is generated, whereas, as we have seen, there is no such content. (Carnap 1932, pp. 78-79)

Ontological questions like 'Are there numbers' may be construed in such a way as to "face an important question." But this is a practical question about which linguistic frameworks we should adopt. The answer 'Yes, there are numbers' or 'No, there are not numbers' is not something that should be viewed as true or false, certainly not as pertaining to what there is in reality. It is only a practical decision that can be more or less useful to the communicative task at hand.

It should be clear that with this argument in "Empiricism, Semantics, and Ontology" Carnap is not endorsing nominalism or nihilism. He is not saying that abstract entities or material objects do not exist. For Carnap, it is just as much of a pseudo-statement for the philosopher to say that material objects or numbers do not exist as it is to say that they do. To focus on the case of numbers, nominalist paraphrases of sentences that look to commit us to abstracta, of the kind for example undertaken by Hartry Field (see Chapter 2), may be motivated. (Although Carnap himself embraced the use of mathematical frameworks.) We might think it is better to replace our current scientific language, which uses mathematical

terms and concepts, with a new scientific language that does not use such terms. But even if the nominalist could do this and preserve our scientific theories, this would not mean he has shown that numbers are not real. He will only have shown something about languages – that we don't need to use a mathematical language to do science.

### **4.3 Responses to Carnap's Arguments**

There are three standard responses that have been given to Carnap's arguments and more generally the logical positivists' critique of metaphysics. It is fair to say, given the vitality of metaphysics today, that most philosophers now find at least one if not all of these responses compelling.

Certainly, the most contentious part of Carnap's arguments both in "Elimination of Metaphysics through the Logical Analysis of Language" and "Empiricism, Semantics, and Ontology" is the verificationist theory of meaning that both arguments presuppose. Let's grant for the sake of argument that at least some metaphysical statements are incapable of being verified by any means whatsoever. Must that entail that these statements are meaningless?

Consider for example some of the statements involved in the realism/nominalism debate. The realist says, "There are universals." The nominalist says, "There are no universals." Is it really the case that these statements are meaningless? A metaphysician could say it is at least clear that these statements contradict each other. One says something does exist; the other says that same thing does not. Mustn't these sentences at least mean something if they are to contradict each other?

We could also point to examples of statements that are uncontroversially meaningless. For example I might say, "Te flob schwubs jip," or "Zee gromple." Surely these are cases of statements that truly are meaningless. Should we really lump "There are universals" into a class with sentences like that?

Moreover, even if we set aside these cases, consider all of the ramifications of the verificationist theory of meaning. There are all sorts of sentences, not just distinctly metaphysical, but also religious and scientific sentences that, if we accept the verificationist theory of meaning, we will be forced to also reject as meaningless. Could we ever verify whether there is a God (in the sense that many people today care about)? Whether there is life after death? Whether there are other worlds causally disconnected from our own (as many current cosmological theories suggest)? Many would argue: it is one thing to say that a claim is meaningful. It is another thing altogether to say it is something we are capable of verifying. To conflate these two issues (semantic and epistemological)



is a confusion. The verificationist theory of meaning has long been out of the mainstream for this reason.

A second critique that has often been offered against these arguments is that they are self-undermining. Carnap wanted arguments that would show the claims of the metaphysicians are meaningless, but his arguments if they work are too powerful; they would show not only that the claims of metaphysicians are meaningless but that his own claims too are meaningless. And if Carnap's own claims are meaningless, how are we to believe what they say? To see this, consider some of Carnap's own claims:

'An alleged statement of the reality of a system of entities is a pseudo-statement without cognitive content' (1950, p. 214), and

'The meaning of a statement lies in the method of its verification' (1932, p. 76).

According to Carnap himself, if such claims are going to be true, then they must be verifiable by either analytic (logical) or synthetic (empirical) means. But how is this supposed to work? These don't look like mere truths of logic or truths capable of verification by any observation we could undertake. These look themselves like philosophical claims to be evaluated along the lines of the philosophical claims Carnap is criticizing. Perhaps Carnap's claims are not intended to be interpreted as internal statements, as statements evaluable within a linguistic system with rules for sentence formation and verification. Perhaps, instead, he intends for us to interpret them only as external statements and themselves meaningless. But then if they are meaningless, what is the point of writing them?

A.J. Ayer (1910-1989), a British philosopher who did much to popularize and defend the views of the logical positivists, has suggested that Carnap might have taken this last position. Wittgenstein, whose *Tractatus Logico-Philosophicus* Carnap and the Vienna Circle much admired, claimed at the end of that work:

My propositions serve as elucidations in the following way: anyone who understands me eventually recognizes them as nonsensical, when he has used them - as steps - to climb up beyond them. (He must, so to speak, throw away the ladder after he has climbed up it.) He must transcend these propositions, and then he will see the world aright. (Wittgenstein 1922, p. 189)

Ayer (1959) suggested that Carnap and the other logical positivists intended their statements as a prescription for how to view philosophy. Philosophy should aim not to make statements that can be verified, but instead to elucidate the statements that can be verified (and those that cannot). Whether this idea of elucidation without meaning is coherent is an issue in itself.

Rather than pursue it, let's instead move on to a third critique that has been made of these arguments, in particular the charge that metaphysics is meaningless and in that way defective with respect to claims that are made in other domains.

In his paper, "Two Dogmas of Empiricism" (1951), aimed directly at Carnap and his circle, Quine aimed to show that the claims of metaphysics are no more problematic than the claims of science, at least in the respects Carnap targeted. We have already discussed the first "dogma" of Quine's title; this is the distinction between analytic and synthetic statements (statements that are verified by exclusively analytic, logical means, and statements that are verified through empirical means). The other "dogma" to which Quine's title refers is a claim of Carnap's that all meaningful claims can ultimately be reduced to claims about perceptual experience. This second claim, the reductionism, was seen above (see Text Box 4.1) in our discussion of protocol statements. For Carnap, all verification and hence all meaning, for the class of synthetic statements, ultimately leads back to what can be directly verified by means of sensory experience.

Quine argued that there is no clear way to make a distinction between statements that are analytic on the one hand and those that are synthetic on the other. One might think that there is an easy separation. For example, in logic or mathematics all claims get verified analytically; in sciences like physics or biology, all claims get verified synthetically, but Quine suggests this isn't so. Here is a famous passage from "Two Dogmas of Empiricism":

The totality of our so-called knowledge or beliefs ... is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements. . . . But the total field is so undetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to re-evaluate in the light of any single contrary experience. No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole.

If this view is right, it is misleading to speak of the empirical content of an individual statement - especially if it be a statement at all remote from the experiential periphery of the field. Furthermore it becomes folly to seek a boundary between synthetic statements, which hold contingently on experience, and analytic statements which hold come what may. *Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system.* Even

a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws. Conversely, by the same token, *no statement is immune to revision*. Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle? (Quine 1951a, pp. 39-40)

In this passage, Quine proposes we view our total system of beliefs as like a web. Each belief is connected to others by various support relations: one belief may give us reason to hold onto others; if we reject that belief, this may give us reason to reject others. And we can see that Quine wants to replace Carnap's picture of a chain of verification ultimately leading back to a basic set of beliefs verified directly by experience with a picture where no belief is ultimate. There is no place where the chain ends. In this web of belief, Quine suggests that some beliefs lie closer to the outside, the periphery; others lie closer to the center. Those at the outside of the web bear more direct connection to our sense experience, our empirical evidence. For example, beliefs like, "The sky is blue," or "The bell is ringing," lie closer to the outside of the web. They are the sorts of beliefs that are most directly confirmed or undermined by what we observe. On the other hand, beliefs like "Triangles have three sides," or to give Quine's example, the law of the excluded middle (for any proposition  $p$ :  $p \vee \neg p$ ), lie closer to the center of the web of belief. These propositions are more insulated from what we observe and are not directly supported or undermined by empirical evidence.

Typically, propositions like that the bell is ringing are held to be synthetic and verified directly by observation. Those like the law of excluded middle are held to be analytic, true by logic alone. However, what Quine wants to emphasize is that this distinction is not real. Even belief in the

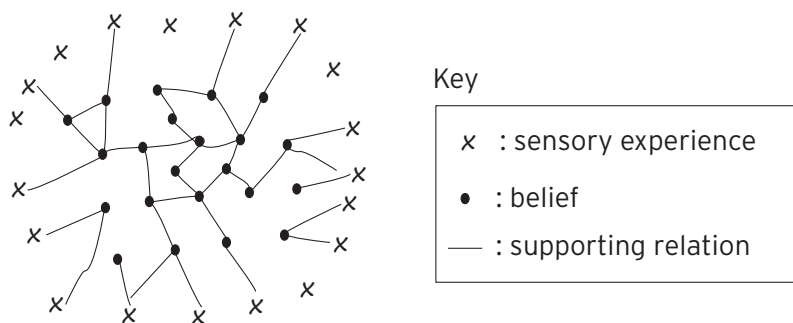


Figure 4.2 Quine's Web of Belief

proposition that the bell is ringing may (reasonably) be rejected in the face of seemingly-confirming empirical observation. For example, your sense experience might involve a certain sound, but you might have reason to believe you are under the spell of a hallucinatory drug. Maybe you are the victim of some kind of massive deception, as in the movies *The Matrix* or *Inception*. In these cases, it may seem to you as if the bell is ringing, but still this may be wrong, something you would wish to reject. Beliefs like this may lie, in Quine's picture, more towards the outside of the web. Sensory experience is more directly connected to beliefs like this than (say) the law of excluded middle. But even the most synthetic-seeming beliefs are not solely ever verified on the basis of empirical observation alone. Even they depend for their confirmation on considerations involving one's overall belief system.

Similarly, those beliefs located closest to the center of the web are the most distant from sensory evidence and might thus be thought incapable of being undermined by empirical observation. The law of excluded middle is something all students learn to prove in their first course in logic. The proof is simple and so this proposition may seem to be an example of an analytic truth if anything is. But as Quine points out, one might on the basis of empirical observation find reason to reject the law of excluded middle. For example, we learn from quantum mechanics that there are states in which particles' features may not be determinate. For example, consider the proposition that a certain electron is in the left-hand side of a box at a particular time  $t$ . According to the law of excluded middle, this proposition must be either true or false: Either the electron is in the left-hand side of the box at  $t$  or it is not in the left-hand side of the box at time  $t$ . But quantum mechanics tells us this is not so. There are certain states electrons can be in, superpositions of position, in which it is neither true that an electron is in the left-hand side of the box, nor is it false. So, says Quine, even a statement one might have thought of as a paradigm of analyticity may be undermined by experience.

What Quine hopes to show using this image of the web of belief is that Carnap is wrong to think that just because in metaphysics our statements are not immediately confirmed by empirical observation, nor are they verified purely as a matter of logic, this means they are defective in some way; that they are meaningless. Quine's point is that nothing we believe is really confirmed solely by empirical observation or by means of logic alone. This is the case even in science.

Ontological questions, under this view, are on a par with questions of natural science. Consider the question whether to countenance classes as entities. This, as I have argued elsewhere, is the question whether to quantify with respect to variables which take classes as values. Now Carnap has maintained that this is a question not of

matters of fact but of choosing a convenient language form, a convenient conceptual scheme or framework for science. With this I agree, but only on the proviso that the same be conceded regarding scientific hypotheses generally. Carnap has recognized that he is able to preserve a double standard for ontological questions and scientific hypotheses only by assuming an absolute distinction between the analytic and the synthetic; and I need not say again that this is a distinction which I reject. (Quine 1951a, p. 43)

Note that here Quine is not rejecting Carnap's view that in answering ontological questions like "Are there classes?" or "Are there material objects?" we must attend to the practical value of adopting such frameworks. What we are doing is not just deciding the answers to these questions using logical or empirical methods. Rather we are, as Carnap put it, settling a pragmatic question, making a practical choice about which overall system of beliefs we want to adopt. Quine accepts all of this, but he suggests that this is always how we decide what we believe. Indeed this is the case in science too. So if ontological claims fall prey to verificationist worries, then so would scientific claims. But scientific claims aren't meaningless or problematic. So, neither are ontological claims.

### EXERCISE 4.2

#### The Web of Belief

Order the following beliefs from 1 to 6, with '1' indicating the belief that is closest to the center of the web.

- The belief that apples have seeds
- The belief that  $1+1 = 2$
- The belief that apples taste sweet
- The belief that apples are fruit
- The belief that everything is either a fruit or not a fruit
- The belief that electrons exist

## 4.4 Present Day Worries about Metaphysical Method

Today, as we saw, the verificationism lying at the core of the logical positivists' critiques of metaphysics is very widely rejected. Many would argue it is simply a confusion to think that for a statement to be meaningful or true, it is necessary that anyone be able to actually verify it. Truth or meaning is one thing, verification is another. Be that as it may, this point doesn't completely eliminate worries related to the verification of metaphysical claims. Rejecting the verificationist theory of meaning, one might

concede that statements about the existence of universals or numbers, or metaphysical views like mereological universalism and nihilism are meaningful, but still worry. Even if these claims are meaningful, still isn't it legitimate to be concerned that they don't seem capable of verification? None of these claims are logical truths. They are substantive claims about the types of entities that exist at our world. And yet at the same time, they seem completely incapable of empirical/synthetic verification. Whether there were universals or numbers or not, whether there were chairs or only simples arranged chairwise, everything would look exactly the same to us. But if we can't verify these claims, then how could we ever know whether they are true or not? Verificationism about truth or meaning may be confused, but verificationism about knowledge is another matter. How could we possibly ever know the truths of metaphysics if we cannot verify them analytically or empirically?

This is the sort of post-positivist worry about metaphysics that is presented in the 2007 book by philosophers of science James Ladyman and Don Ross *Every Thing Must Go*. In this book, Ladyman and Ross argue that while science is capable of making true *and* justified claims about the way the world is, metaphysics, at least insofar as it proceeds detached from science, is not. Ladyman and Ross, although they explicitly associate themselves with Carnap and the logical positivists, do not want to reject all metaphysics as meaningless because they reject the verificationist theory of meaning. They view metaphysical claims as meaningful but worry that the methods metaphysicians have used for coming to discover truth are problematic.

Their main worry about much of contemporary metaphysics is that instead of basing their arguments in scientific statements that possess empirical justification, metaphysicians argue for their positions using a priori intuitions about how the world must be or what seems right to them. Ladyman and Ross argue that it is implausible to think that nature has endowed us with any special way of uncovering objective truth by intuition alone. It is not possible to discover what the world is like fundamentally just by reflecting on what seems right. Indeed, our intuitions have been proven again and again over the history of ideas to get things wrong. To counter this reliance on intuitions, Ladyman and Ross propose a new prescription for how to do metaphysics in such a way that it can discover objective truths about the world: "metaphysics naturalized."

The proposal is that metaphysicians follow what they call the **Principle of Naturalistic Closure**:

Any new metaphysical claim that is to be taken seriously at time *t* should be motivated by, and only by, the service it would perform, if true, in showing how two or more specific scientific hypotheses, at least one of which is drawn from fundamental physics, jointly explain

**Principle of Naturalistic**

**Closure:** the principle that any metaphysical claim to be taken seriously at a time should be motivated by the service it would perform in showing how two or more scientific hypotheses, at least one of which is drawn from fundamental physics, jointly explain more than what is explained by the hypotheses taken separately

more than the sum of what is explained by the two hypotheses taken separately. (Ladyman and Ross 2007, p. 37)

This is an extremely strict rule in that it would severely limit the sorts of claims metaphysicians could make, if they are to be viewed seriously. Metaphysics would be limited to the task of merely unifying claims made in the sciences. For example, one might show the constitutive relation between the states described by sciences like biology or psychology and the atoms and molecules of physics.<sup>2</sup> This may be an interesting task, but is it a specifically metaphysical one? And are these the only sorts of issues we can hope to settle in metaphysics? Ladyman and Ross think that this is the only way for the claims of metaphysics to have legitimacy and avoid the use of intuitions – by drawing on a basic set of information that has been confirmed by the methods of science. Today, many metaphysicians happily embrace the label of being “naturalistic metaphysicians,” metaphysicians who use science to inform their metaphysics. But this doesn’t mean they view their task as merely linking together claims made in different sciences.

#### 4.5 The Relationship Between Metaphysics and Science: A Proposal

Ladyman and Ross’s critique of metaphysics has stirred up a lively debate about the proper methodology for metaphysics and its relation to science. Since the crux of their critique focuses on the methods metaphysicians use to justify their claims, most replies have focused on clarifying and defending the methodology of current metaphysics. For example, Laurie Paul, in a 2012 article “Metaphysics as Modeling: The Handmaiden’s Tale” (alluding ironically to the frequent remark that philosophy is the “handmaiden of the sciences”) argues like Quine that the methodology of metaphysics is really not that different from the methodology of science itself. So, if one wants to be critical of the methods used by metaphysicians to establish their claims, then one must be critical as well of the methods scientists use to establish theirs.

It is worth noting that unlike Quine however, Paul does not think this commits us to a merely pragmatic view of metaphysics: that metaphysics just gives us a body of claims it would be useful to accept. Just as we ordinarily think scientific claims may be confirmed and are true (rather than merely useful for one purpose or other), so too may those of metaphysics be confirmed and true. The question is how in metaphysics and in science this truth gets discovered.

Paul argues that in both metaphysics and science, the central task of the theorist is to construct **models**. Models are theoretical structures involving a set of basic representational devices (new terms and concepts,

**Model:** a theoretical structure involving a basic set of representational devices accounting for a set of data

and principles involving these concepts) that can account for a set of data including but not limited to those observations we may make. In both science and metaphysics, different models may adequately represent the domain under consideration. For example, in physics, a theorist may use a model using the concept of a field to represent the kinds of entities there are (as in electromagnetic theory). Another theorist may use a model that involves only the representation of particles. Similarly in metaphysics, one theorist may use a model that represents the world in terms of the concepts of particular and universal. Another may use a model that rejects the concept of a universal and conceives everything as a more or less complex collection of tropes.

Paul emphasizes that it is an oversimplification to see science as confirming the statements that make up its models simply on the basis of empirical verification.<sup>3</sup> Empirical methods may take us as far as a certain point. But there will always be rival models equally capable of explaining the empirical data. The scientist must supplement empirical criteria with the use of theoretical criteria like simplicity, unificatory power, internal and external consistency, and fruitfulness (ability to generate further hypotheses) to decide between models.<sup>4</sup> Thus when the philosopher looks at rival metaphysical theories and assesses them on the basis of criteria like which is simpler, which is consistent with other things she believes, and so on, she is not applying a kind of methodology that is different from that used by scientists themselves.

Paul also argues that intuitions are used just as much in science as in metaphysics in order to construct and evaluate models. A critic might ask what can be learned from thinking about the Ship of Theseus or the case of Lump and Goliath, but cases like this are used in science just as much as in metaphysics. There are a series of famous thought experiments used by Isaac Newton to justify his basic physical principles.<sup>5</sup> The founders of quantum theory similarly used a series of thought experiments in order to justify their theories, the most famous of which is Schrödinger's cat. And we will see in Chapter 8 the sort of thought experiment Einstein used to argue for his special theory of relativity.

**Thought experiments** are more or less detailed descriptions of fictional cases from which the scientist or philosopher draws out consequences in order to assess a given theory. Since these cases are fictional, these consequences are not discovered on the basis of observation. In science and metaphysics, empirical data is of course used to construct and evaluate models. If a model contradicts what we observe, then this is a reason to throw it out. But since the empirical data only constrains the choice of models up to a point, there is a significant role for a priori methods in both science and metaphysics. This involves, first, the evaluation of rival models according to their success at accommodating the theoretical virtues and, second, the use of thought experiments. According to Paul,

**Thought experiment:**  
a fictional case used to draw out consequences of use to the building and assessment of a scientific or philosophical theory



the main difference between science and metaphysics does not lie in a difference in methodology.

This is not to say that there is no difference between metaphysics and science. According to many, metaphysics engages a kind of question more fundamental than the kind of question addressed in science. And so there is a difference in subject matter. For example, in physics, we may ask what kind of space or spacetime there is at our world: one that is flat and Euclidean, or one that is curved and non-Euclidean. In metaphysics, we may ask whether there *is* such an entity as space or spacetime at all or whether all that exists are relations between material objects. In physics, one might ask what kinds of basic particles there are. In metaphysics, we may ask what is the best way to think of these particles, as concrete entities instantiating universals or as bundles of tropes. Thus, Paul and others have argued, it seems that the difference between the disciplines of metaphysics and science consists in their subject matter, not their methodology. There is no significant difference in methodology.

Paul's proposal for how to understand the relationship between metaphysics and science is not the only possible way to see metaphysics as constituting a legitimate research program. However, her proposal certainly reflects the way many contemporary metaphysicians view what they are doing. One controversy that remains is whether metaphysics indeed aims at a more fundamental set of truths than those discovered by science (as Paul argues) or whether metaphysics and science (and perhaps religion as well, if we reject naturalism) are better seen as operating jointly to discover what are a common set of fundamental truths. This is a subtle issue on which we will remain neutral for much of this text.

### EXERCISE 4.3

#### The Handmaiden to the Sciences

What do you think it would mean for metaphysics to be the handmaiden to the sciences? Do you think this is a good metaphor for metaphysics, something metaphysics should aspire to? Or do you agree with Paul that metaphysics has a larger role to play in our understanding of the world? If so, what is that role?

### Suggestions for Further Reading

In addition to the papers of Carnap cited earlier, one can find more classic works of the logical positivists in the volume *Logical Positivism* edited by A.J. Ayer. Ayer's own *Language, Truth, and Logic* was a very influential popularization and development of the views of the logical positivists as

well. In addition to Quine's critique of the logical positivists in his "Two Dogmas of Empiricism," his paper "On Carnap's Views on Ontology" is also worth a look. Matti Eklund, Huw Price, and Scott Soames's articles in the volume *Metametaphysics*, edited by Chalmers, Manley, and Wasserman, also provide good overviews of the dispute between Carnap and Quine.

In addition to the contemporary critique of metaphysics of Ladyman and Ross, other neo-Carnapian critiques have recently been given by Mark Balaguer (2021), Eli Hirsch (2010), Thomas Hofweber in the *Metaphysics* volume, and Amie Thomasson (2015). John Hawthorne and Theodore Sider respond to Hirsch's critique in their contributions to the *Metametaphysics* volume.

The July 2012 issue of the journal *Philosophical Studies* was a special issue devoted to the relationship between metaphysics and science and contained many articles on that topic in addition to the one by Laurie Paul discussed above. See especially James Ladyman's "Science, Metaphysics, and Method," which responds to Paul's proposal. For another view about the relationship between metaphysics and science, see Steven French and Kerry McKenzie's article, "Thinking Outside the Toolbox." My paper "Are the Questions of Metaphysics More Fundamental Than Those of Science?" challenges Paul's proposal by examining several conceptions of relative fundamentality.

## Notes

- 1 We will discuss the topic of essence in Chapter 10 on modality.
- a Though see Susanna Siegel (2010) for a rival view.
- 2 Though would even statements about atoms and molecules constitute parts of *fundamental* physics? Note how strict Ladyman and Ross's rule is.
- 3 Here, especially, one may note strong similarities between Paul's critique of Ladyman and Ross and Quine's critique of Carnap and the logical positivists.
- 4 For discussion of these virtues, see Kuhn (1977).
- 5 See Sklar (1974) for an overview of some of these.

# 5 Natural and Social Kinds

## Learning Points

- Introduces traditional essentialism about natural kinds
- Introduces the homeostatic property cluster view of natural kinds
- Presents an influential account of how kinds may be socially constructed and yet, at the same time, real
- Introduces the distinction between descriptive and revisionary metaphysics, as well as the ameliorative approach to ontology
- Considers how debates about social ontology may feed back to inform approaches to ontology more generally

## 5.1 A Traditional Account of Natural Kinds

In Chapter 3, we discussed several puzzles that arise when one asks whether in addition to the most basic things (mereological atoms or simples), there exist in addition complex things built out of them. It is common to think these puzzles may be solved, and that in addition to the most simple things described by our fundamental physical theories, there exist also a variety of mereologically complex objects: from molecules and cells, to bacteria, birds, and human beings; various sorts of inanimate objects: stars and planets, mountain chains and galaxies; and human artifacts: chairs, statues, and typewriters. In this chapter, we will ask what makes something not a mere collection of atoms or simples, but an object of a particular kind, such as an electron as opposed to a tiger. A distinction that will be important in this chapter and the next will be the distinction between natural kinds and social kinds. After introducing these concepts and views about what makes something a natural or a social kind, we will be able to discuss a topic ubiquitous in academic debates about what it means to say that something is socially constructed.

In his story "The Analytical Language of John Wilkins," Jorge Luis Borges (1899–1986) describes an encyclopedia in which:

animals are divided into (a) those that belong to the emperor; (b) embalmed ones; (c) those that are trained; (d) suckling pigs; (e) mermaids; (f) fabulous ones; (g) stray dogs; (h) those that are included in this classification; (i) those that tremble as if they were mad; (j) innumerable ones; (k) those drawn with a very fine camel-hair brush; (l) etcetera; (m) those that have just broken the flower vase; (n) those that at a distance resemble flies.

It is clear that such a division is in many respects different from the classificatory scheme used by biologists. One using Borges's encyclopedia would divide the animals up in a way represented by Figure 5.1. The biologist would divide animals up in a way represented by Figure 5.2. Although the former classification is certainly interesting, many would argue that the biologist's classification is metaphysically privileged in some way in that it somehow tracks the objective structure of the world. But in what way is the biological classification so privileged?

Many, even most, metaphysicians believe the world has an objective structure. To use Plato's metaphor, there are certain classifications that are privileged in the sense that they "carve nature at its joints."<sup>1</sup> It is the aim of both metaphysics and science not just to find the list of objects there are but also to find these joints in nature, the real classifications of objects. These are what metaphysicians refer to as the **natural kinds**. Natural kinds aren't merely important for descriptive purposes, however. They also play an important predictive role. As Quine pointed out in his 1969 essay "Natural Kinds," when we know that an object is a member of a natural kind, we are able to reliably predict its future behavior. For example, if you know the liquid in the glass in front of you is water, you can reliably predict what it will do to you if you drink it. And if you know that instead it is vodka, you can reliably predict it will do something very different to you if you drink it.

**Natural kind:** a classification of objects that corresponds to objective joints in nature

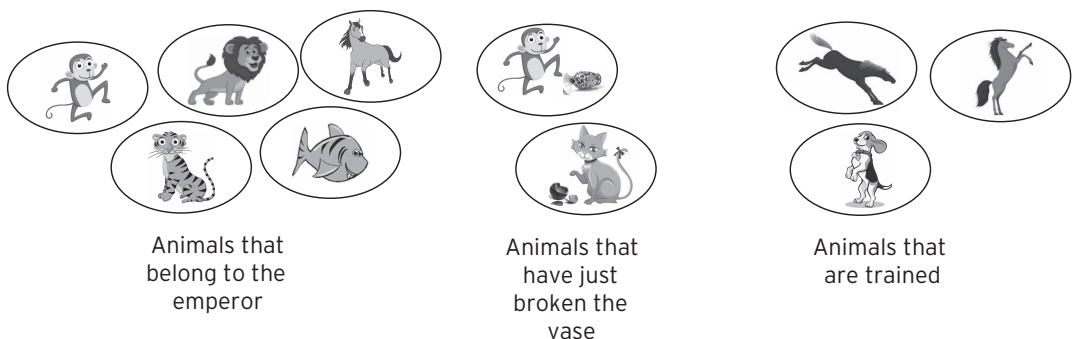


Figure 5.1 Borges' System of Classification



Figure 5.2 Biological System of Classification

According to many philosophers, water ( $H_2O$ ) and other chemical compounds are paradigm examples of natural kinds. They are kinds of substances that have intrinsic essences. And these essences are important because they underwrite stable patterns of cause and effect. You drink water, you get hydrated, which makes your body function better. You drink alcohol, you get dehydrated, and your body functions less optimally. Quine and other philosophers of science called this **projectability**. Natural kinds are **projectible** in the sense that when you are able to subsume some object under a natural kind, you are able to reliably predict (or project) what you will observe in the future.

**Projectible:** a kind is projectible if you can reliably predict what things of that kind will do in future circumstances

There are several ways in which scientists divide the world into natural kinds: the classification of animals given by biologists is one. Biologists divide the animals of the world using a sequence of finer and finer grained classifications (phylum, class, order, family, genus, species). Other paradigm examples of natural kinds are the chemists' classifications of substances into elements (hydrogen, helium, lithium . . .) ordered by atomic number in the Periodic Table, as well as into chemical compounds ( $H_2O$ ,  $NaCl$ , etc.), and the physicist's classification of elementary particles (quarks, electrons, neutrinos, and so on) summarized in the Standard Model of particle physics.

So, what is required for a group of entities to form a natural kind? One traditional view is that natural kinds are groups of objects that share a common **essence**. That is, each member of the natural kind has some features that make it the kind of thing that it is.<sup>2</sup> So for example, the natural kind *human being* is a group of entities that includes me, you, Elon Musk, Aung San Suu Kyi, and so on. According to this view, all of us are members of the kind in virtue of sharing some common essence. Perhaps the essence of a human being is to have a certain genetic profile. Plato considered the view that it was to be a featherless biped. The correct answer is for science to discover. However that turns out, according to the traditional view, the essence of human beings will consist in a set of intrinsic features we humans all share that make us part of the natural kind *human being*. There are no human beings that lack this essence. And there are no entities that fail to be human beings that have it.

**Essence:** a set of features that make something the kind of thing it is, or the particular object that it is

This traditional essentialist view about natural kinds is thought to apply as well to the case of chemical elements. For example, anything that is a member of the natural kind *gold* (Au) will have atomic number 79, that is, it will consist of atoms with exactly 79 protons each. Anything that is a member of the natural kind *oxygen* (O) will have atomic number 8, that is, it will consist of atoms with exactly 8 protons each. It turns out that some chemical elements don't seem to occur "naturally," at least not on Earth. These are the synthetic elements, for example, *einsteinium* (Es, atomic number 99), which was discovered following the explosion of the first hydrogen bomb. Despite the fact that these elements are manmade, in the sense of being caused to exist by human technologies, einsteinium

and the other synthetic elements are still natural kinds in the sense that what it is for something to be a member of one of these kinds is just the possession of a certain atomic number. The element's essence does not involve any human factors.

Those metaphysicians who believe in natural kinds take these kinds to reflect objective groupings in the world. Entities who are members of a common natural kind are genuinely similar in some way in all possessing a common essence (genetic profile, atomic number, set of physical features, etc.). The fact that these entities are similar in the relevant way is not subjective or relative to our own contingent classifications of the objects. The objects in question would bear these similarities whether we thought to group them in these ways or not.

### EXERCISE 5.1

#### The Traditional View of Natural Kinds

- A. Metaphysicians who adopt the traditional view typically take water ( $\text{H}_2\text{O}$ ) to be a paradigmatic example of a natural kind. Which of the following properties constitute the essence of water? Which do not?
- Being composed of molecules with two hydrogen atoms
  - Being flavorless
  - Being a liquid
  - Being a liquid at room temperature ( $20^\circ\text{C}$ )

For each answer, explain why or why not.

- B. Explain in your own words why einsteinium can still count as a natural kind, even though it was created by human technologies. What is another example of something that is a natural kind, but created by human technologies?

### TEXT BOX 5.1

#### Essences, Universals, and Nominalism

It might be helpful to connect this discussion of natural kinds with our previous discussion in Chapter 2 about properties. We have seen that a common view about natural kinds says that objects are members of a natural kind in virtue of being objectively and intrinsically similar to each other in some way. If one is a realist about universals, then one would understand this as saying that members of a natural kind all instantiate some common set of universals. Katherine Hawley (1971-2021) and Alexander Bird argued for a view along these lines, that natural

kinds are complex universals (universals built up out of other more basic universals) (2011). However, it should be emphasized that having an essence requires more than just instantiating some universals. For an object's essence is thought to make it the kind of object that it is. And not any set of universals makes for an essence. For example, if mass of one kilogram and positive charge are universals, then we might consider the group of all objects that have mass of one kilogram and positive charge. This would be a group of genuinely similar objects, but it isn't a natural kind. Being positively charged and having mass of one kilogram doesn't make something a distinctive kind of object.

One might ask whether one could believe in natural kinds if one rejects universals and is instead a nominalist. The answer is "yes," but it depends on what kind of nominalist one is. To believe in natural kinds, one must believe there are some objective classifications and similarities in nature. For example, one might think there is something metaphysically special about the group of electrons in that the electrons are all intrinsically similar in some way, whereas there fails to be anything metaphysically special about the group of objects in my desk. These objects are not intrinsically similar. As David Lewis showed in his paper "New Work for a Theory of Universals," there are a couple of ways one could do this while rejecting the existence of universals. First, one could think there are basic facts about similarities between certain objects. The electrons are all similar in some way, while the objects in my desk are not. This is something that those who like a trope theory of properties could accept. The objects in a natural kind that possess certain tropes but not others have certain similarities to each other in virtue of instantiating tropes that are similar. Trope theorists will generally think it is just a brute fact about nature that certain tropes are similar to each other, while others are not.

Another way of making sense of natural kinds while rejecting universals would be to adopt some form of class nominalism, and then argue that there are certain classes that are privileged in virtue of being the ones corresponding to objective similarities in nature. These are what Lewis calls the natural properties (discussed in Chapter 2 Section 2.5). For example, the class containing all and only the negatively charged objects is a natural property, but the class containing all and only the objects in my desk is not. If we want to think of natural kinds too as sets or classes of objects, then the natural kinds would be a special group of natural properties, the ones in which all members share a common essence.

## 5.2 The Homeostatic Property Cluster Account

We have just described the traditional view about natural kinds, that they are groups (or sets) of entities that share a common essence. This view entails that there are properties each member of a kind has. These may be understood as providing a set of necessary and sufficient conditions for membership in the kind. In addition, advocates of the traditional view have often assumed the essences shared by members of a natural kind characterize these entities intrinsically, specifying ways they are in themselves.

In the 1990s, the philosopher of science Richard Boyd (1942-2021) challenged this traditional view of natural kinds, arguing that at least some



paradigmatic natural kinds, the biological species, do not satisfy the traditional account. If we consider any biological species, it is just not true that there is any set of intrinsic properties that each and every member of this species (and only the members of this species) has in common. First, there is variation. And second, what seems to make something a member of a given species isn't an intrinsic fact about it, but a matter of its history, including the organism's evolutionary lineage. And so, Boyd argued, if we want to make sense of the idea that biological species are natural kinds, we will have to move beyond the traditional account.

Boyd's proposed account of natural kinds has become very influential in philosophy of science and is becoming more widespread in metaphysics, especially in discussions of social ontology. It is called the **homeostatic property cluster (or HPC) account of natural kinds**. According to this view, the members of a natural kind need not share a common set of intrinsic properties. Instead, for there to be a natural kind, it is sufficient that there exist families of properties that are stably clustered in nature, where the stability of these clusters is due to the presence of what Boyd calls a **homeostatic mechanism**. This is a mechanism that ensures rough stability over time in such a way that "either the presence of some of the properties in the family tends ... to favor the presence of the others, or there are underlying mechanisms or processes which tend to maintain the presence of the properties in the family, or both" (1999). Boyd notes that for this to make for the sharing of a common kind, it must be that the stable clustering of a family of properties makes for the instantiation of theoretically or practically interesting causal patterns. This is important if we want to have an account that accommodates Quine's point mentioned above, that natural kinds are projectible.

Boyd argues that it is uncontroversial that biological species satisfy the HPC account of natural kinds, noting:

A variety of homeostatic mechanisms - gene exchange between certain populations and reproductive isolation from others, effects of common selective factors, co-adapted gene complexes and other limitations on heritable variation, developmental constraints, the effects of the organism-caused features of evolutionary niches, etc. - act to establish the patterns of evolutionary stasis which we recognize as manifestations of biological species. (1999)

The core point is that there is a host of factors that make it the case that members of a common species stably share certain families of properties. This does not mean that every member of the species is exactly similar in each possessing the same intrinsic features, but it does mean that members will overlap in many of the properties they have (these may include both intrinsic and extrinsic properties) and this will be stable over time.

**Homeostatic property cluster account of natural kinds:** the view according to which members of a natural kind need not share a common set of intrinsic properties; rather, for there to be a natural kind, it is sufficient that there exist families of properties that are stably clustered in nature, where the stability of these clusters is due to the presence of a homeostatic mechanism

**Homeostatic mechanism:** a mechanism that ensures stability over time in such a way that either the presence of some properties tend to favor the presence of the others, or there are underlying mechanisms or processes which tend to maintain the presence of the properties, or both

Thus, species satisfy the HPC account of natural kinds, even if they do not satisfy the traditional account.

### **EXERCISE 5.2**

#### **The HPC Account of Natural Kinds**

Select a biological species and list five different properties that members of this species typically share. These are properties that would form part of what Boyd calls the homeostatic property cluster. Are the properties you listed all intrinsic properties? If not, which of them are extrinsic properties?

There are two interesting consequences of Boyd's account in addition to what has already been mentioned. First, there may be cases in which it is indeterminate whether a given object is a member of a natural kind. According to the traditional account, kind membership is an on/off matter. Either an object has the essence or it does not. On Boyd's account, all that is required to be a member of a kind is that something have enough of a family of properties – but there is no rule for what counts as “enough.” This is, according to Boyd, as it should be. In the case of biology, and probably other taxonomies as well, there will be indeterminacy or vagueness as to whether something is a member of that natural kind.<sup>3</sup>

Second, which natural kinds there are is something that can change over time as different homeostatic mechanisms emerge. This again is to be expected, and marks an improvement over the traditional account, if one thinks there is not a permanent fact about which biological and other kinds exist in nature, but rather that which kinds there are is something that evolves over time.

One final thing to note about the HPC account of natural kinds is that, as Boyd emphasizes, it is compatible with the view that natural kinds are associated with essences. It is just that on the HPC account, essences won't involve a single set of necessary and sufficient conditions that every member of a natural kind must meet, nor will the essences necessarily be constituted by intrinsic features of the kind's members. The essences on an HPC account may include extrinsic features as well.

### **5.3 Social Kinds and Social Reality**

The HPC account of natural kinds seems promising if we want a more permissive account of natural kinds than the traditional essentialist account.

It does not seem, however, that all of the kinds of things that exist ought to be classified as *natural* kinds. Yes, there are objective classifications given to us by the natural sciences: the classifications of physicists, chemists, and biologists. But there are more classifications as well. For example, there are classifications of objects into artifacts of various sorts (tables, pencils, laptops, money, artworks) and classifications of people in various ways (wives, attorneys, students, refugees, Buddhists, Republicans, curmudgeons).

It seems some of these classifications track real distinctions in the world, and yet metaphysicians would generally say that such classifications do not correspond to natural kinds. This is for two reasons. First, these are not kinds that are studied by the natural sciences. Rather, to the extent that these classifications of people and objects are studied scientifically, they will typically be the subject of one or another of the social sciences: anthropology, sociology, economics, or political science.

Second, natural kinds are supposed to track what is objective structure in nature. And there is a sense of what is meant by 'objective structure' according to which these classifications are not objective. This is the sense in which what is objective is mind-independent. Without minds, there wouldn't be any laptops, attorneys, or Buddhists. So, there is at least this one sense in which the kinds above fail to be objective.

This isn't to say, however, that there is no sense in which these classifications are objective. Another sense of 'objectivity' ties it to **intersubjectivity**. In this sense, what is objective is what may be agreed upon by different people (or subjects). There are certain classifications that are clearly subjective, for example, what counts as good music. One person likes country music. Another can't stand it. Here we think there isn't an objective fact of the matter about who is right. And there is little hope for intersubjective agreement, one set of facts about what counts as "the good music" that could be agreed upon by everyone. On the other hand, facts about what is a pencil do seem to be objective in this intersubjective sense. Everyone can agree about what is a pencil and what is not. These facts are not subjective like the facts about good music.

So, it seems like there are kinds that are not natural kinds, but still are kinds with objective membership criteria, in this second sense of the term. And these kinds that we have been discussing, kinds that seem brought into existence by the result of some human behavior or interactions, are what metaphysicians refer to as **social kinds**. In the next section, we will discuss one common way of explaining how it is that social kinds come into existence, that is, by social construction. As we will see, this term has a variety of connotations in the metaphysics and social science literature.

**Intersubjectivity:**  
allowing for agreement  
by different people or  
subjects

**Social kinds:** kinds  
that depend for their  
existence on human  
behavior or interactions

We will be interested in the ways of conceiving of social construction that relate it to matters of ontology.

## 5.4 Social Construction

As Ian Hacking emphasizes in his book *The Social Construction of What?*, often when theorists argue for a social constructionism thesis, they are not making a claim about the construction of a metaphysical category or kind; rather they are making a claim about the construction of a concept, or set of concepts. And so often, their interest is not primarily metaphysical. For example, one might be interested to argue for a social constructionism about morality, or more specifically the view that our concepts of good and evil are socially constructed. This is something that (famously) Friedrich Nietzsche (1844-1900) did in much of his work.<sup>4</sup> The idea here is to show that there isn't some mind-independent, necessary fact about which actions are good and which are evil, but rather that these concepts are human constructions. They were made up by people:

Whatever has *value* in our world now does not have value in itself, according to its nature – nature is always value-less – but has been *given* value at some time, as a present – and it was *we* who gave and bestowed it. Only we have created the world *that concerns man!* (Nietzsche, *The Gay Science*, p. 301)

According to Nietzsche, morality and the concepts of good and evil were made up in order to make sure the weaker people in society could be controlled and wouldn't mind sacrificing themselves for the wills of the more powerful. Nietzsche believed that we as a society should reevaluate these moral concepts that were constructed in the past and replace them with a set of concepts that would better allow human flourishing.

Hacking argues that in general, when one puts forward a social constructionism thesis about some concept X (e.g., morality), one thinks that the following facts about X obtain (1999, pp. 6-12):

- (0) In the present state of affairs, X is taken for granted; X appears to be inevitable.
- (1) X need not have existed, or need not be at all as it is. X, or X as it is at present, is not determined by the nature of things; it is not inevitable.

And very often they also think:

- (2) X is quite bad as it is.
- (3) We would be much better off if X were done away with, or at least radically transformed.

Constructionism thus implies a debunking of assumptions previously held about X. We may have previously thought that the concepts of morality we have are inevitable and they couldn't be otherwise, but the social constructionist argues actually that is not true. These concepts have a contingent history and we as a society could (and perhaps should) replace them with other concepts.

As Hacking notes, there are many examples of social categories for which these four theses fail to apply. For example, the list of examples of social kinds above included the category *attorney*. Few would argue that this category is inevitable, and so, in this case, thesis (O) seems not to obtain. We know our world could have existed without attorneys, that the existence of attorneys depends on people interacting in certain sorts of ways, and that the facts about who is and is not an attorney depend on a variety of contingently existing social structures and rules. For Hacking, when someone presents a thesis of social construction, one makes this claim against a background in which people generally do not recognize that the category is the result of contingent histories and social patterns. The philosopher Ron Mallon expresses this point by saying that the social kinds generally of interest in academic discussions of social construction are *covert*, by which he means "the existence, persistence, or specific properties of the category are believed to be the product of natural facts, rather than human decision, culture, or social practices" (2016, p. 49). This is in contrast with social categories that are *overt*, those for which most already believe the existence and persistence of the category are the result of contingent histories and social practices, e.g., categories like attorneys, legal spouses, or cryptocurrency.

So, let's suppose we adopt a social constructionism about moral concepts that involves Hacking's four theses. We can note that by itself, this doesn't yet commit us to any *metaphysical* thesis about the construction of social kinds. This is because although we would then view the concepts of good and evil as contingent and socially constructed, ontological questions about the existence of any corresponding kinds would still be left open. For example, we could decide that the right thing, ontologically speaking, is to say that moral categories don't exist. There never was such a thing as good and evil – we just acted as if there was, and this was a mistake. This would be an eliminativist position. But another ontological position compatible with Hacking's four theses is that by constructing the concepts or categories, we also created new kinds. This would be to adopt a kind of realism about morality. If we accept theses (2) and (3), we would also think the fact that such kinds exist is not a good thing; they are something we should try to eliminate. The main metaphysical puzzle for the social constructionist, according to Mallon, is to figure out the circumstances in which the correct ontological position is realism (rather than eliminativism) about the kinds. It isn't in general the case that our

representations or concepts cause a new kind to come into existence, so how does this work in the cases where kinds do come to exist? This is the subject of the next section.

## 5.5 Constructing Human Kinds

In this section, we will discuss three responses to this metaphysical challenge of when the social construction of concepts brings with it the social construction of new kinds. These are responses to the challenge that have been proposed by three philosophers: Hacking, Mallon, and Åsta. Note that all three of these philosophers are particularly interested in the social construction of *human* kinds, kinds that we as human beings may or may not fall under. This is not to say that there is not a lot of discussion in the metaphysics literature about the construction of other social kinds. For example, there is a good amount of discussion on the social construction of money, corporations, and social groups. Social ontology is a flourishing area of research, and it's especially been picking up in the past decade. The reading list at the end of the chapter directs the reader to some good starting points on work concerning the construction of social kinds other than human kinds.

The first account of the construction of human kinds is Hacking's. This account rests on a thesis about human action famously put forward by the philosopher Elizabeth Anscombe (1919–2001) in her 1957 book *Intention*. This is that the same bodily movement can constitute many different actions according to the intention under which it is undertaken. In one of her examples, a man rhythmically moves his arm pumping water into a cistern that supplies drinking water to a house. In this case, we might ask: what is the man doing? Anscombe notes, there could be many legitimate answers to this question: he is pumping water, earning wages, supporting a family, beating out a curious rhythm to amuse himself, replenishing the water-supply, pumping poison into the water-supply to kill the fascist party chiefs who live inside, overthrowing the government, instituting the Kingdom of Heaven on Earth, and securing a good life for all the people (1957, p. 37). Anscombe argues that the correct answer to which action he is performing depends on the man's personal intentions. And some of these intentions involve how he categorizes himself: as a husband, a father, an activist, a Christian.

Hacking follows Anscombe in noting that how one categorizes oneself can affect the intentions one has and thus which kinds of actions one can perform. But he also notes, in turn, by enacting different kinds of intentional actions, one can become a particular kind of person. For example, I think of myself as an activist. This causes me to engage in certain kinds of behavior. In turn, my engaging in this behavior makes me an activist. For this reason, he calls some human kinds **interactive kinds**. These are

**Interactive kind:** a kind that is the result of and in turn causes certain intentional actions

kinds that are the result of and in turn cause certain intentional actions. Moreover, Hacking emphasizes that the interactive kinds in which he is interested don't simply result from an isolated person thinking of themselves and their actions in certain special ways, but also from a network of people's perceptions and treatment of each other:

People, including children, are agents, they act, as the philosophers say, under descriptions. The courses of action they choose, and indeed their ways of being, are by no means independent of the available descriptions under which they act. Likewise, we experience ourselves in the world as being persons of various kinds ... The awareness of what kind of person one is may be personal, but more commonly is an awareness shared and developed within a group of people, embedded in practices and institutions which are assigned in virtue of the way in which they are classified ... We are especially concerned with classifications that, when known by people or by those around them, and put to work in institutions, change the ways in which individuals experience themselves - and may even lead people to evolve their feelings and behavior in part because they are so classified. Such kinds (of people and their behavior) are interactive kinds. (1999, pp. 104-105)

Where we find interactive kinds, there are generally looping effects, Hacking notes. My conceiving myself as falling under a particular category allows me to perform certain actions and this in turn strengthens my and others' understanding of myself as falling under the category in question, which then causes me to perform more actions and so the loop continues (1999, pp. 109-110).

How does this answer our central metaphysical question, of in which cases the construction of concepts leads to the construction of metaphysical kinds? Consider the fact that there may be ways we have of categorizing ourselves or others that do not manifest themselves in stable patterns of action over time. These classifications, as such, don't make a difference in the world; they don't change the world, and, in this sense, they seem to be nothing more than classifications. However, when how we categorize ourselves changes how we behave, how we are seen and treated by others, and thus changes further actions I and others perform, we have an interactive kind. Category concepts that denote interactive kinds track genuine causal patterns in the world, and it is in virtue of this fact that the categories in question are real.

Mallon (2016) raises two objections to Hacking's account of the construction of human kinds as interactive kinds. First, Mallon argues that Hacking's account requires for a kind to be socially constructed, we must come to the category as the result of an interactive social process. But

Mallon cites research in cognitive science suggesting that many categories for human kinds are innate. Humans form these concepts as the result of a natural, developmental process, not as the result of social interactions. For example, Mallon cites scientific research suggesting that humans are evolutionarily predisposed to classify themselves and others in terms of categories of in-groups and out-groups. And yet many metaphysicians argue that in-groups and out-groups are socially constructed categories.

Mallon's second objection to Hacking's account of kind construction concerns that account's emphasis on how individuals self-identify and their conscious intentions in acting under those self-identifications. Mallon wants to allow that this is often the case with human kinds, but in some cases, one will fail to consciously identify under a category one is a member of. Sally Haslanger has also used this point to criticize Hacking's account, providing the example of widowhood. As Haslanger notes, one need not identify as a widow in order to be a genuine member of this category (2012, p. 128).

Both of these objections to Hacking suggest that an account of the social construction of human kinds should not require that their construction involve explicit self-identifications or social interactions. As Mallon notes, both unintentional or automatic mental processes, as well as physical and environmental changes, can also play a central role in the construction of human kinds.

Mallon's 2016 proposal is that the construction of human kinds requires first the emergence of what he calls *social roles*. A social role exists when:

- (SR1) There is a term, label or mental representation that picks out a category of persons *C*, and that representation is associated with a set of beliefs and evaluations (that is, a conception) of the persons so picked out, and
- (SR2) Many of the beliefs and evaluations in the conceptions of the role are common knowledge in the community.<sup>5</sup>

Mallon lists several kinds of beliefs and evaluations that may figure in the "conception" of a category of persons. These include:

- Beliefs about names or labels used for the categories, e.g., attorneys are in some places called 'lawyers' or 'barristers'
- Beliefs about conditions of ascription, e.g., if one has passed the bar exam and received a license to practice law, then one is an attorney
- Beliefs about essential features, e.g., necessarily, attorneys are rational
- Beliefs about category-typical features, e.g., attorneys are typically well-paid, or typically skilled at deciphering complicated documents



- Evaluations, e.g., attorneys are smart, or attorneys shouldn't be trusted
- Norms, e.g., when interacting with an attorney, one should be careful what one says. (2016, p. 59)

When a conception is established as common knowledge in a community, then, Mallon argues, this establishes a social role in that community.

The existence of a social role does not on its own support realism about a corresponding kind. Mallon notes that the social roles that are of interest to metaphysicians are those that do not just exist, but are **entrenched**. This is to say that they track causal structures that are significant and so may support predictions (“Oh she is an attorney, thus she will probably understand this complicated contract I’ve been asked to sign”) and explanations (“Oh she is an attorney, that explains why she is able to afford this Tesla Roadster”). It is the social roles that become entrenched that metaphysicians would argue track genuine distinctions in nature.

Here, Mallon is in agreement with Hacking, who also argued that socially constructed human kinds are causally potent. However, as noted earlier, this need not in all cases involve humans identifying under a category and then acting intentionally under this description. Mallon argues there are (at least) three different pathways through which the conception of a category can causally differentiate the members of that category and create social roles that are entrenched: (i) through intentional actions by and toward category members (as Hacking had suggested), (ii) by shaping the manifestation of automatic (unintentional) cognitive processes by and toward category members (e.g., through implicit bias or stereotype threat), and (iii) by guiding behaviors that shape the environment (in the form of legal institutions, cultural institutions, norms used to regulate people’s behavior, modifications to the physical environment such as the construction of neighborhoods, hospitals, statues, sidewalks, and ramps). In Mallon’s work, he provides a rich discussion of all three sorts of causal pathways and examples of how they may turn shared conceptions of human categories (social roles) into efficacious social kinds.

Mallon argues that we can understand the way in which the covert and entrenched social roles he discusses count as genuine human kinds by way of Boyd’s homeostatic property cluster account of kinds, discussed in Section 5.2. Of course, this does not mean the kinds in question are thereby natural kinds. Rather, they are social kinds. They are kinds whose existence depends on the behavior of humans, humans who represent themselves and each other in certain stable ways, reinforcing these representations through both conscious and unconscious cognitive processes, and the shaping of environments. But nonetheless, as these social processes result in homeostatic processes, processes that make certain

**Entrenchment:** a social role is entrenched when it tracks causal structures that are significant and so may support predictions and explanations.

properties tend to be instantiated stably in clusters over time, the result is social roles that are kinds in the sense Boyd described.

A third and distinct account of the social construction of human kinds we will consider is the conferralist account proposed by Åsta in her 2018 book *Categories We Live By*. Åsta has similar concerns to those raised by Haslanger and Mallon about Hacking's proposal, in particular, noting that the social construction of human kinds need not operate by way of one's intentionally acting under a particular identification. Like Mallon, she concedes that this happens some of the time, but that one's falling under a social category at times comes about not because of one own's intentions, but because of those of others. This is why she calls her account '**conferralism**.' According to Åsta, the social construction of human categories arises from some in a community (those with social standing) *conferring* certain kinds of status onto others. Typically, this status is conferred on a member of a community in virtue of their being perceived by others to instantiate a certain underlying feature, what Åsta calls the base feature. The official statement of this account relies on the notion of social significance:

For a feature B, the base feature, to have **social significance** in a context is for another feature F, the constructed feature, to be conferred upon people taken to have B. (Åsta, 2018, p. 44)

Conferralism is the view that a kind (or feature) F is socially constructed when its base feature B has social significance in a context. This gives the person taken to have B a certain social status, that of being a member of the kind F. So, if being an attorney is our constructed feature F, then this account will imply that there is some other feature B, perhaps having passed the bar exam and possessing a license to practice law, that has social significance in a context (e.g., the state of California). When a person is taken to have passed the bar exam and possess a license to practice law in the relevant context, this makes them a member of the human kind attorney.

Importantly, Åsta notes that to say that someone has a certain social status is to say that they have certain constraints or enablements on their behavior. When one is taken to be an attorney, one has the ability to do certain things that others are not permitted to do. Or, to use an example from Haslanger of another obviously socially constructed category, the "cool kids" can do things the uncool kids cannot. They can get away with things in the contexts in which the status is conferred that the uncool kids cannot. Having a social status makes a difference to how one can or cannot behave. The uncool kids have certain constraints on their behavior, where the cool kids have enablements. And so again with this account, we see those categories that are socially constructed being those that track causal structures. It is important to emphasize Åsta's language in the

**Conferralism:** the view that a kind (or feature) F is socially constructed when its base feature B has social significance in a context

**Social significance:** a feature B has social significance in a context when another feature F is conferred upon people taken to have B

above statement of her account. For a human category to be conferred in this way, it is not necessary that a person actually have the base feature in question, only that they are “taken to have” the base feature.

Àsta illustrates her account using the example of disability. It will be helpful to work through this example in order to illustrate some of the account’s key concepts. Àsta argues in her book that disabilities are examples of socially constructed human categories.<sup>6</sup> For disability to be a socially constructed category in Àsta’s sense, there will generally be a base feature B that people with social standing in a certain context take others to have. B will be a feature that has social significance in the context, so that those taken to have B gain certain constraints or enablements in their behavior. In the case of disability, Àsta proposes the base feature is physical impairment. Individuals are taken to have certain kinds of physical impairments. This forms the basis for the social construction of disability, when these physical impairments become socially significant in a context, thus granting people who are taken to have them a social status (disabled) that “consists in constraints on and enablements to their behavior.” For example, she notes there are:

institutional constraints, such as laws banning people with physical impairments from driving, even though the people in question could physically drive . . . There can also be enablements. People taken to have physical impairments can be granted certain allowances, such as being permitted to park in certain designated areas, or to preside over religious rituals. (2018, pp. 44-45)

Àsta argues that most if not all of the human categories we use to classify ourselves and others are socially constructed in this sense, of being social statuses that are conferred upon us. In the next chapter, we will discuss how she uses this account to formulate a view about gender.

The differences between Àsta’s view and the previous two positions are subtle. According to all three accounts, the social construction of human kinds is a process whereby people represent themselves or others in ways that lead to changes in the world, changes in the behaviors of those categorized and/or others, and, as Mallon especially emphasizes, changes to the physical environment. We can see that Àsta, like Mallon, and unlike Hacking, does not want to require for one to fall under such a category that one oneself identify under that category. The conferralist view rather makes it that the conferrals of social categories may be made by any of those with standing in one’s community. Àsta also distances herself from the part of Mallon’s account requiring that the construction of a human category comes with a “conception” of the category that becomes common knowledge in the community. On Àsta’s view, members of a community may start using a classification to track an underlying feature without

having a robust conception of what those with that feature are like or even that they are using the classification to enforce social patterns in the community (2018, p. 53). So long as that feature is, as a matter of fact, a source of social significance in the community and leads to enablements and constraints by those taken to have it, then social construction will take place, according to Åsta's conferralism. These differences between the views can help us when we want to consider a particular social category and ask which account of construction seems most correct in that case.

### EXERCISE 5.3

#### Three Accounts of Social Construction

Of the three accounts discussed in this section, which do you take to be the best account of the following human kinds. Or, is it wrong in the first place to say that kind is socially constructed? Explain your answer.

- a. Being a Buddhist
- b. Being a parent
- c. Being healthy

## 5.6 Causal vs. Constitutive Construction

It is important to emphasize that Mallon and Åsta's accounts of the social construction of kinds are not primarily intended to provide *causal* accounts of how social kinds come to be. Rather they are interested in providing *constitutive* accounts, accounts of what it is for something to be a socially constructed kind. It is neither necessary nor sufficient for a kind to be socially constructed that it be caused by social processes.<sup>7</sup> For example, some earthquakes may be caused by social processes, by business decisions that lead to hydraulic fracking. But even if some earthquakes are human-caused, this doesn't make earthquakes a human kind. What it is to be an earthquake is just to be a certain kind of geological process. Moreover, as Mallon pointed out, some of the categories we use to classify other humans may be innate. If so, then our use of such representations is not caused by social processes. This nonetheless does not rule out these categories' being human kinds, kinds that individuals belong to in virtue of their place in a pattern of human interaction.

It is of course interesting in many cases to understand that categories and distinctions we thought were natural are actually the historical result of social processes. This shows us how our classifications of each

other are not inevitable, and as we saw Hacking emphasize, this is often an important point for those who want to argue that certain categories are socially constructed. But this sense of construction as a historical process is not what is the primary focus of the metaphysical accounts of Mallon and Åsta. These accounts show us not merely how social categories *come to be*, but *what they are*. For Åsta, the nature of social kinds involves statuses that are conferred on others by those with social standing in a particular context. For Mallon, the nature of social kinds involves conceptions of others that become common knowledge and entrench themselves in certain robust ways, including in conscious and unconscious mental attitudes and changes to the environment.

## 5.7 What Are We Doing When We Do Social Ontology?

Hopefully the last couple of sections have helped to make it clearer what is typically meant when a metaphysician makes the claim that a particular human category is a social rather than a natural kind, and what it means to say that a particular kind is socially constructed, but for all that, real. In the next chapter, we will look at metaphysical debates concerning two varieties of human categories: races and genders. And we will ask three questions about these categories. First, are they real or should we be eliminativists about them? Second, if they are real, are they natural kinds or social kinds? And, finally, if they are natural (or social) kinds, then what sort of natural (or social) kinds are they? Before we get there, however, it will be useful to first turn to issues of methodology, asking what is the correct method to use when approaching questions about human kinds, whether they exist, and what they are.

In his book *Individuals* from 1959, P.F. Strawson (1919–2006) introduced what is now a famous distinction in metaphysics. He said:

Metaphysics has been often revisionary, and less often descriptive.

**Descriptive metaphysics** is content to describe the actual structure of our thought about the world, **revisionary metaphysics** is concerned to produce a better structure. (1959, p. 9)

Descriptive metaphysics is often thought of as conceptual analysis. We humans have certain concepts, certain ways of thinking about the world. And it is natural for us to want to understand these concepts or ways of thinking better, to better understand how we think about e.g., free will or causation (as we'll get to in later chapters), or material objects (as we discussed in an earlier one). But in metaphysics, we often want to do something more than just ask about how we actually conceptualize these phenomena. We want to understand whether these ways we have of thinking actually correspond to anything real, anything that exists.

**Descriptive metaphysics:** a metaphysical project of describing the actual structure of human thought about the world

**Revisionary metaphysics:** a metaphysical project of improving the structure of human thought about the world

To see this, let us return to an example that occupied us throughout Chapter 1, that of Pegasus. Here the project of descriptive metaphysics would involve looking to our ways of thinking about Pegasus, asking what is involved in the concept. Here the answer is a winged horse, a very special winged horse, the one that, according to Greek myth, was captured by Bellerophon. Asking is there anything real, anything that exists that is this winged horse, the answer is “no.” And so we have this concept of Pegasus, this way of thinking about a horse that we can describe and investigate. But there is no such entity: Pegasus, the winged horse.

Here, though, we saw Quine’s character McX make a move. McX suggested that perhaps instead of thinking of Pegasus in the way we typically do, as a horse, we instead should come to think of Pegasus as a mental object, as a concept or idea. We can all agree that the concept or idea of a winged horse exists. So, McX proposes, if we *revise our understanding* of Pegasus so that it no longer refers to a horse, but instead to a mental entity, then we can correctly say that Pegasus exists. And this may be a good thing, if we can now have a solution to Plato’s problem about non-existent objects.

Most metaphysicians today would not be tempted to accept this move, to revise our understanding of Pegasus from that of a nonexistent physical object to that of an existent mental one. However, the project of revisionary metaphysics, as Strawson described it, is very much alive and well. We will see it again and again throughout this text, in discussions of social constructionist theories of race and gender, the B-theory of time, four-dimensionalist theories of persistence, regularity theories of causation, and compatibilist theories of free will. Strawson says that the aim of revisionary metaphysics is “to produce a better conceptual structure.” Here, it will be useful to pause for a moment and ask what is meant here by ‘better.’

One legitimate sense of what is meant by a “better” conceptual structure is one that is **epistemically** better, that is, one that is more correct, or better represents the objective structure of reality – hence, one that gives us more knowledge. This is certainly one thing in which the revisionary metaphysician is interested. The revisionary metaphysician wants to develop and defend representations of our world that accurately represent the objective structures there are.

**Epistemic:** related to knowledge and truth

However, there is also a broader sense in which one might take a conceptual structure to be “better.” This is a sense in which a conceptual structure might not just help those who adopt it epistemically, in forming more correct beliefs about the objective structure of the world, in having more knowledge about the nature of reality, but might help those who adopt it more broadly, in promoting not only knowledge and understanding, but also practical aims in achieving personal goals, or aims with objective value, such as promoting a more ethical, just, and beautiful world.

**Ameliorative project:** a metaphysical project that starts from the questions of what is the point of having a certain concept and what concept would do this work best

Haslanger distinguishes those projects in metaphysics that aim to capture what kinds there actually are in nature – seeking out epistemically better conceptual frameworks – with what she calls **ameliorative projects**:

*Ameliorative projects ... begin by asking: What is the point of having the concept in question – for example, why do we have a concept of knowledge or a concept of belief? What concept (if any) would do the work best? ... If we allow that our everyday vocabularies serve both cognitive epistemic and practical purposes that might be well-served by our theorizing, then those pursuing an ameliorative approach might reasonably represent themselves as providing an account of our concept – or perhaps the concept we are reaching for – by enhancing our conceptual resources to serve our (critically examined) purposes. (2012, pp. 367–368)*

For some cases, like the puzzle about nonexistent objects, we might question what purposes (beyond the epistemic) would be served by revising the meaning of terms like ‘Pegasus’ so that they refer to mental rather than physical entities. However, for many debates in social ontology, the concepts we use matter. And there is reason to believe that a revision in our conceptual schemes might sometimes be a good way to promote values we have beyond better tracking what there is. In these cases, the metaphysician may explicitly adopt the ameliorative approach. They will not merely evaluate a proposed ontology based on the extent to which it accurately depicts reality, but engage as well with the question of what practical purposes such an ontology may serve. And this includes the extent to which an ontology might contribute or detract from our pursuit of ethical or political ideals. Although the ameliorative approach was introduced by Haslanger in the context of a debate about social ontology, its use and significance is much wider. This is a good example of how recent debates in social ontology are feeding back to produce general lessons for metaphysics more broadly.

### **Suggestions for Further Reading**

On natural kinds, a classic source is W.V. Quine’s “Natural Kinds.” Richard Boyd develops his homeostatic property cluster account in a series of works including his “Homeostasis, Species, and Higher Taxa.” Also see the essays collected in the volume *Carving Nature at its Joints* edited by Joseph Keim Campbell, Michael O’Rourke, and Matthew H. Slater, and *Scientific Enquiry and Natural Kinds: From Planets to Mallards*, by P.D. Magnus. Gideon Rosen’s “Objectivity and Modern Idealism: What is the Question?” very helpfully distinguishes several different conceptions of objectivity. The books *Objectivity* by Lorraine Daston and Peter Galison and *The View from Nowhere* by Thomas Nagel are also recommended.

A good place to start to get a broader view on the main issues in social ontology, especially ontological issues about the kinds found in the social sciences, is Brian Epstein's book *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. Muhammad Ali Khalidi's article "Three Kinds of Social Kinds" is a nice introduction to some of the metaphysical issues surrounding social kinds and their distinction (or not) from natural kinds. John Searle's books *Making the Social World* and *The Construction of Social Reality* develop an account of the construction of social facts. See "All Illuminating Exchange: *The Construction of Social Reality*," by Barry Smith and John Searle, and Amie Thomasson's "Foundation for a Social Ontology" for critical engagements with Searle's work. An excellent book about the social construction of disability is *The Minority Body*, by Elizabeth Barnes.

On the ameliorative approach, see the essays in Haslanger's book *Resisting Reality*, as well as the essays in the volume *Conceptual Engineering and Conceptual Ethics*, edited by Burgess, Cappelen, and Plunkett, especially the essay by Esa Díaz León, "Descriptive vs. Ameliorative Projects." See also Amie Thomasson's "What Can We Do, When We Do Metaphysics?"

## Notes

- 1 Plato, *Phaedrus*.
- 2 We will return in Chapter 10 to discuss further the concept of essence and essentialism.
- 3 This is a point we discussed in Chapter 3.
- 4 For example, in *The Genealogy of Morals*.
- 5 This is a paraphrase of the account stated at Mallon (2016), p. 58. 'Common knowledge' is a technical term that comes from David Lewis (1969). Something is common knowledge when everyone knows it, everyone knows everyone else knows it, everyone knows everyone else knows everyone knows it, and so on.
- 6 Social constructionism is a common view about the metaphysics of disability, though it comes in several forms. See Barnes (2016) for a useful overview.
- 7 This is a point emphasized both by Brian Epstein in his book *The Ant Trap*, Section 2.3, and by Sally Haslanger in her *Resisting Reality*, pp. 87ff. We discuss constitution or "grounding" theses in more detail in Chapter 7.



# 6 The Metaphysics of Race and Gender

## Learning Points

- Introduces the three main views on the metaphysics of race and gender: biological realism, social constructionism, and eliminativism
- Examines why social constructionism is often regarded as the correct view about both topics in philosophy
- Compares different forms of social constructionism
- Discusses the difficulties surrounding essentialism about race and gender

## 6.1 Three Views about the Metaphysics of Race and Gender

In this chapter, we will focus on two topics in social ontology: the metaphysics of race and gender. There are three main views that appear in discussions of these human categories: biological realism, social constructionism, and eliminativism. **Biological realists** believe that races (or genders) are biological kinds, such that one's membership in that kind depends on one's instantiating certain biological features. **Social constructionists** believe that races (or genders) are social kinds, such that one's membership in that kind is constituted by some existing social practices or institutions. **Eliminativism** is the view that races (or genders) fail to exist, and that our practices of classifying individuals in terms of races (or genders) are the result of a false theory and should be discontinued.

Although at one time, biological realism was the dominant position about both race and gender in philosophy and the wider culture, today, philosophers largely reject this position due to its lack of scientific support. Instead, the dominant position is social constructionism. In this chapter, we will discuss this move to social constructionism in more detail

### Biological realism about

**Xs:** the view that Xs are biological kinds

### Social constructionism

**about Xs:** the view that Xs are social kinds

### Eliminativism about

**Xs:** the view that Xs are not real

and what the position looks like when applied to race and gender, as well as the challenges constructionist accounts face. We will also discuss why philosophers have largely tended to reject eliminativism about both race and gender, even though there are some reasons to find eliminativism attractive.

In this chapter, we will discuss these various positions first as they apply to race, and then as they apply to gender. Although we will see some of the same kinds of arguments arising in the two cases, race and gender are different human categories and so should be given separate consideration. We should not automatically assume that arguments that are compelling when applied to one group of categories will be sound when applied to the other.

As we will see in what follows, today's work on the metaphysics of race and gender is often conducted using the ameliorative approach to metaphysics introduced in Chapter 5. The metaphysician adopting an ameliorative approach to these topics will ask not only which natural kinds and social practices exist to ground facts about race and gender, but also:

What is the point of having concepts for the various races and genders?

Which accounts of these concepts would best serve our epistemic and practical goals?

More specifically, metaphysicians seeking ameliorative accounts of race and gender are typically guided by an anti-racist, feminist perspective. There is variation in what is meant by 'anti-racist' and 'feminist' among philosophers, however a useful definition is provided by bell hooks (1952-2021), who understands **feminism** as:

a commitment to eradicating the ideology of domination that permeates Western culture on various levels - sex, race, and class, to name a few - and a commitment to reorganizing ... society so that the self-development of people can take precedence over imperialism, economic expansion, and material desires. (1982, p. 195)

We will work with this definition in mind going forward.

## 6.2 Biological Realism about Race

In biology, we find various classifications of living things, including into kingdoms (plants, animals, fungi), classes (mammals, reptiles, birds), and species (humans, wolves, red jungle fowl). There are also subspecies of many species. For example, the domestic chicken is a subspecies of the red jungle fowl (*Gallus gallus*). Each of these classifications is a natural kind where, following the traditional conception of natural kinds introduced in Chapter 5, it is thought that the members of that group share

**Feminism:** a commitment to eradicating the ideology of domination that permeates Western culture on various levels - sex, race, and class, to name a few - and a commitment to reorganizing society so that the self-development of people can take precedence over imperialism, economic expansion, and material desires

a common biological essence. This essence may be identified as a phenotype, meaning it describes some observable characteristics common to members of that category, or as a genotype, in terms of the types of genes the members of that category carry. To say, as the biological realist does, that races are biological kinds is to say that the members of a given race all share a common **phenotype** (set of observable features) and/or **genotype** (set of genetic features). For some biological realists, the different races are subspecies of the species *Homo sapiens*.

Biological realists will often concede that it is not possible to individuate the races based on phenotype alone. Although there are broad similarities people associate with different races including skin color, eye shape, and hair texture, there is no such combination of observable features that is essential to making a person a member of a given race (Zack 2017). People may often pass for members of a race other than their own. Thus, biological realists will turn to genotypic characteristics to individuate the races.

Biological realism thus seems to require establishing that there is a particular genetic profile shared by all White people that is different from that shared by all Black people that is different from that shared by all Asian people and so on. However, many argue, this fact does not seem supported by scientific results. As Anthony Appiah notes (1992, p. 35), the scientific consensus for some time has been that there is greater genetic variation among members of a given race than there is between the races themselves (Nei and Roychoudhury 1972). This has been confirmed by geneticists sequencing the human genome, leading scientists at the U.S. National Institute of Health to announce at the White House in June 2000 that there is no genetic basis for a distinction among the races, that there is only “the human race.”

In defense of their view, biological realists today tend to point to the migration patterns of our ancestors, noting that millions of years ago, different groups of people settled in distinct regions of the Earth. Robin Andreasen, for example, argues that races are “ancestor-descendent sequences of breeding populations, or groups of such sequences, that share a common origin” (1998, p. 200). Andreasen thus grounds the existence of different races in terms of biologically recognized facts about the histories of human populations, rather than in terms of specific genetic profiles. Similarly, Quayshawn Spencer has argued that we find groups “anchored in a region circumscribed by major geographic barriers to human interbreeding,” e.g., oceans, mountain chains, and deserts (2019, p. 99). Spencer counts five such human continental populations, which correspond to the five races the U.S. government uses on its Census forms and other official documents (at least as of 2020): Whites (Eurasians), Blacks (Africans), Asians (East Asians), Native Americans, and Pacific Islanders (Oceanians). Furthermore, although earlier scientific results

**Phenotype:** a set of observable features

**Genotype:** a set of genetic features

would seem to support the claim that there is no genetic basis for race, Spencer argues, citing the work of the biologist Noah Rosenberg and his collaborators (2002), that newer research into the human genome does reveal sufficient difference between individuals in these five groups to support biological realism, even if these differences have not been found to have any behavioral consequences. This conclusion is controversial, but continues to be debated among biologists and philosophers.

It is important to emphasize one difference between forms of biological realism that were influential in the past and those that are held by scientists and philosophers today, like Andreasen and Spencer. Historically, biological realists have been committed to what Appiah calls **racialism**. This is:

the view . . . that there are heritable characteristics, possessed by members of our species, which allow us to divide them into a small set of races, in such a way that all the members of these races share certain traits and tendencies with each other that they do not share with members of any other race. (1992, p. 13)

Many of those who have maintained biological realism about race in the past believed that members of a given race, in virtue of having the biological features essential to being members of that race, are as such, determined to behave in certain ways, be better or worse at certain activities, or have certain intellectual capacities and emotional dispositions that others lack. In addition, many of the people who maintained (or maintain) biological realism were (or are) racists – they saw (or see) some races as being superior or inferior to others, in virtue of these purported behavioral or psychological capacities. Serious defenders of biological realism today, like Andreasen and Spencer, will emphasize that their view does not rest on the assumptions of racialism and racism. Although they will argue that the races can be distinguished in terms of genetic differences between individuals, scientific consensus does not support the claim that the genetic differences that make for racial classifications in turn determine the emotional or behavioral dispositions or intellectual capacities of individuals. This is not to say that biological differences among the races do not give rise to any salient differences between individuals, beyond those at the genetic level. Spencer (2018) considers the possibility that the biological features that make for racial classifications could determine the risks different individuals have for developing certain diseases. But this certainly does not commit him to the racialism of earlier views.

So, if biological realism can be separated from racialism and racism, and we allow that scientists are still debating whether it is reasonable to think that in our evolutionary histories, human populations settled in different, isolated regions in such a way so as to cause there today to be

**Racialism:** the view that there are heritable characteristics which divide human beings into a small set of races, in such a way that all the members of these races share certain culturally significant traits and tendencies with each other that they do not share with members of any other race

different groups of people with different genetic characteristics, then why are most philosophers not biological realists today?

**Mismatch objection:** the objection to biological theories of race that there appears to be a mismatch between our racial categories and the human continental populations

The first reason is that biological realism seems to face what Joshua Glasgow calls the **mismatch objection**. This is that there appears to be a mismatch between our racial categories and the human continental populations. According to Spencer's version of biological realism, as we mentioned, there are five human continental populations, and so five races. On this view, one's race corresponds to which human continental population one's ancestors descend from. Since many people will have ancestors from multiple populations, many people will count as mixed race. One version of the mismatch objection is that since all humans have ancestors from Africa, it seems to follow from this version of biological realism that all humans are Black. But it is not the case that all humans are Black. Therefore, this form of biological realism seems to be incorrect.

There is a fix the biological realist may offer here: this is to refine their view in such a way that one's membership in a given race is tied to facts about one's relatively recent ancestors only. However, even then, critics will note that there are other mismatches between racial categories and the biological categories picked out by the biological realist. For example, as Glasgow notes, racial classifications (in the United States and United Kingdom, for example) count South Asians as Asians, not White. And yet South Asians are descended from the continental populations of Eurasia, not East Asia. As such, Spencer's account entails that South Asians are White, not Asian. But South Asians are Asian. In addition, of the five human continental populations, none corresponds to Middle Eastern or Hispanic. And yet, those who identify as Middle Eastern or Hispanic will often consider these to be distinct races.<sup>1</sup> Many who debate the metaphysics of race, Spencer and Glasgow included, think it is important that a metaphysics of race track the categories of race recognized in one's home context. Thus, as Americans, Glasgow and Spencer are debating race as a phenomenon recognized in the U.S.<sup>2</sup> And thus, the categories those who live in the U.S. recognize and use to classify themselves and each other are legitimate data to be included in assessing Spencer's biological realism as a metaphysical position.

## EXERCISE 6.1

### The Mismatch Objection

Which version of the mismatch objection is the most compelling to you? State the objection in numbered premise form. How might a biological realist like Spencer or Andreassen respond to this objection? Which premise would they reject and why?

This brings us to a second influential objection against biological realism: the **relativity objection**. This objection starts from the premise that if the races are natural kinds, then they are absolute. This premise can be motivated by considering less controversial examples of natural kinds: the biological species or chemical elements. Start with the latter. In the case of the chemical elements (hydrogen, helium, and the rest), we think there is one absolutely correct periodic table, not a plurality of equally correct periodic tables, one that is correct for one culture or nation, another that is correct for another. Similarly, for the biological species, there is just one correct system of classification. If a biologist in one part of the world discovers a new species, this affects the classification system of biologists everywhere. However, systems of racial classification do not seem to be like this: it seems that there are a plurality of equally correct systems recognized in different countries and by different cultures.

Spencer notes that the U.S. Office of Management and Budget, that creates the U.S. Census and other government documents, recognizes five racial categories (White, Black, Asian, American Indian, and Pacific Islander). However, censuses conducted in other countries reflect a different set of races than those given by the U.S. Office of Management and Budget. The Brazilian Institute of Geography and Statistics recognizes five (different) races: Branco (White), Pardo (Brown), Preto (Black), Amarelo (Yellow), and Indígena (Indigenous). In 2011, South Africa recognized four races: Black African, Colored (i.e. "mixed race"), Indian or Asian, and White. The Indian census fails to track racial classifications altogether, at least under that label.

So, there are many different systems of racial classification. But again, if races are natural kinds, then one must be the correct system of racial classification: the one that corresponds to the biological classifications there are in nature. Since different countries' racial categories differ, this would seem to indicate that one is right and the others are wrong. And yet, it does not seem that there could be any basis for claiming that one is more correct than any other. So, it seems racial classifications are not absolute, and so the races are not natural kinds. Biological realism is not the correct view about race. Or so one may argue.

**Relativity objection:** the objection to biological theories of race that different cultures have different racial categories, and yet if races are natural kinds, then they must not vary with culture

## EXERCISE 6.2

### The Relativity Objection

State the relativity objection to biological realism in numbered premise form. What is the best response a biological realist could give to this objection?

### 6.3 Social Constructionism or Eliminativism?

Suppose one grants these last points. The social constructionist will then argue that as racial classifications vary with national and perhaps cultural context, this suggests races are not natural or biological kinds, but rather social phenomena. The existence of stable patterns of racial classification suggests that races should be thought of as social kinds, with different groups of social kinds corresponding to the different racial classifications we find in different nations and cultures. These include the U.S. races, the Brazilian races, the South African races, and so on. In the next section, we will consider a couple of different versions of social constructionism and the challenges that these views face. However, before taking social constructionism on board, let's first ask why it has seemed reasonable to many to continue to endorse some version of realism about race, even if it is a social constructionist realism and not a biological version of realism.

One might think that it would be more reasonable for those who reject biological realism to endorse eliminativism about race. After all, some eliminativists argue, consider the harm that has been and continues to be done when we classify people on the basis of race. Philosophers of race today acknowledge that the introduction of racial concepts was motivated by racist thinking and a desire to dominate and exploit groups seen as "other."<sup>3</sup> Once we learn that there aren't biological distinctions to support this racist thinking, one might think we should recognize that race itself is unreal and discard such classifications.

Often, eliminativists draw on an analogy. Consider the fact that from about 1450-1750 many women in Europe and the Americas were thought to be witches.<sup>4</sup> This practice, of labeling certain people as witches, persists in some parts of the world today. This classification was and is used to justify punishment, banishment, and murder. But this is a mistaken classification. In the sense these women were or are thought to be witches, there are no witches. Today, we recognize this fact, and thus the correct metaphysical position about witches is eliminativism. An eliminativist about race can argue that we should think an analogous view is the correct one to hold about race. In the case of race too, the classifications were introduced in order to justify the oppression of groups of people. But there are no such differences among people that would justify such treatment. And so, we should think, just as there are no witches, there are no White people or Black people, Asians, American Indians, or Pacific Islanders. We should thus stop the use of such classificatory schemes.

This kind of reasoning has had cultural purchase in the U.S., especially in the 1990s when it was thought that the best way to promote equity and justice was to be "color blind" and avoid classifying people on the basis of race altogether. However, there are two influential arguments against eliminativism, and so, although the view is still taken seriously by some philosophers (Appiah (1992, Chapter 2) and Glasgow (2009)), it is not the dominant position.

First, the realists about race will point out that there are differences between members of different races concerning how members of these races are treated, how they view themselves, and material barriers they face, even if there turn out not to be biological differences between them. Social scientists have documented how a person's race affects their life in a variety of ways from the kind of neighborhood one lives in, the kind of healthcare one receives, the educational opportunities one has available to them, to the likelihood of being hired for work.<sup>5</sup> Which race one is also affects one's sense of oneself and one's identity. These effects may be positive, in the sense in which one takes pride in one's racial identity, or negative, as in the case of stereotype threat. This is all to say that races are causally potent: they have genuine (and significant) effects. Metaphysicians will often argue that if something is causally potent, then this implies it real. This is what metaphysicians call **the Eleatic principle**, after a passage in Plato's dialogue *The Sophist*. So, critics of eliminativism argue we should think that races are real.

**Eleatic principle:** the principle that the entities that are real are those that are causally potent

Noting the causal impact an individual's race has on their social treatment, in an influential essay from 2000, "But What Are You *Really*?" Charles Mills (1951–2021) argued that the system of classifying people into races (in particular, in the U.S.) is a vertical system of classification. The classification of people into races is a vertical hierarchy in the sense that members of certain races have a kind of privileged status and power that has enabled them to possess advantages over those who are members of the other races. Mills argues this fact that race is a vertical classification system makes it possible for us to say there are facts about which race one is really, and that gives the question of one's race, in Mills's phrase "ontological depth" (2000, p. 42). If one's race didn't have significant causal impacts, then there wouldn't be an answer to whether one is *really* a member of one race rather than another. There would only be the labels, with nothing real behind them.

### EXERCISE 6.3

#### The Eleatic Principle

Just as realists about race will often use the Eleatic principle to argue *against* eliminativism about race, philosophers will sometimes use the Eleatic principle to argue *for* eliminativism about a certain class of entities. For example, a nominalist (see Chapter 2) might use the Eleatic principle to argue against the existence of universals or mathematical entities. Or a mereological nihilist (see Chapter 3) might use the Eleatic principle to argue against the existence of composite material objects. How might one of these arguments go? What are the premises and what is the conclusion? Evaluate the argument you've stated.



This is the first argument against eliminativism, that race matters in that it has significant causal potency. As such, we should recognize racial categories as real. A second argument against eliminativism is political and comes when we adopt the ameliorative approach to the metaphysics of race described in Section 6.1. If we want to identify and overcome racial injustice, we must first recognize that there are groups of people that are the subjects of racial injustice, and this means recognizing that there are races. For example, the Black Lives Matter movement, founded in 2013 after the acquittal of George Zimmerman in the killing of Black teenager Trayvon Martin, has been highly effective in mobilizing communities to protest the killing of Black people. Especially in the summer of 2020, after the killing of George Floyd by a police officer, the movement successfully organized hundreds of community protests united around this message, and raised millions of dollars in funds to support, *inter alia*, changes in policing, government, and education. But how does one say “Black Lives Matter” and capture the force of that statement and movement if one adopts eliminativism about race, claiming there are no Black people? It seems that to successfully promote racial justice requires some form of realism about the races.

#### **6.4 Working Out a Social Constructionism about Race**

These are the reasons philosophers of race today tend not to be eliminativists. But, if one is to adopt a realism about race and one does not accept biological realism, then one faces the question: what kinds of things are races? The most common view today is that races are social kinds: kinds that depend for their existence on social practices and institutions. Racial categories are ultimately grounded in the way we represent ourselves and others, the way we treat each other, and the kinds of institutions we have created in the form of neighborhoods, schools, and cultural institutions.

The most influential kind of social constructionism about race understands racial categories in terms of social and political structures of privilege and oppression. As Mills (2000) argues, racial classifications seem not to be grounded in actual similarities and differences in evolutionary histories as biological realism would have it. But racial classifications do seem to track the evolutionary histories people are thought by others to have, based on inferences from their appearance. The existence of races then seems to come about when groups of people classified according to these supposed evolutionary histories are then subject to privilege or oppression on this basis. Here is a formal account along these lines proposed by Sally Haslanger:

A group is *racialized* iff its members are socially positioned as subordinate or privileged along some dimension (economic, political, legal,

social, etc.), and the group is “marked” as a target for this treatment by observed or imagined bodily features presumed to be evidence of ancestral links to a certain geographical region.... In other words, races are those groups demarcated by the geographical associations accompanying perceived body type, when those associations take on evaluative significance concerning how members of the group should be viewed and treated. (2012, pp. 236-238)

This is a version of social constructionism because it has the existence of the races grounded in social facts - people’s perceptions and thoughts about other people, and their treatment of each other on this basis. For Haslanger, just as for Mills, in pointing out the reality of race, one is pointing out the reality of a vertical system of structural privilege and oppression, in order to identify, challenge, and overcome it. For Haslanger, this metaphysical account of race is explicitly motivated by the ameliorative project. It is not only an account that tracks features of the world that have significant causal potency, but also one that can be helpful in promoting social justice, by allowing us to speak meaningfully about and effect change regarding race and racism. This is something we could not do effectively if we discarded these concepts altogether.

Haslanger’s account allows us to say meaningfully and truthfully that there are Black people, White people, Asians, American Indians, and Pacific Islanders today (and the races of other countries outside of the U.S.) in order to promote social justice. But some have asked, if Haslanger’s view about what the races are is correct, then what would happen if we were able to achieve an end to racism and structural injustice. This seems to many to be a coherent possibility, even if progress is slow and it is not clear how we might achieve this aim. It seems to follow from Haslanger’s account, however, that if there was no more racism, no more subordination or privileging of groups of people based on their perceived evolutionary histories, then there would no longer be any races. But is this the correct result?

Thinking about this possibility has led philosophers to question the tenability of accounts like Haslanger’s. There are two types of challenges. The first is that it has seemed to many to be simply incorrect that if there were no more racism and race-based injustice, then there would cease to be White people, Black people, Asians, American Indians, and Pacific Islanders. As Glasgow puts it, “Even if tomorrow all groups currently recognized as racial had equal power ... I do not think we’d say that on the ordinary concept of race Hillary Clinton somehow loses her whiteness or that Jeremy Lin stops being Asian because of *those* points of equality” (2019, p. 251).

A related challenge, raised more in the spirit of the ameliorative approach, asks: if there were an end to race-based injustice, would it then

be best for us to eradicate our concepts of race? Chike Jeffers (2019) has proposed that even if race-based injustice were to end, we would still have use for racial distinctions. Jeffers doesn't dispute that racial classifications initially came about because of racist ideology and white supremacy. But he resists the idea that structures of privilege and oppression are the sole basis for our racial classifications today.

Instead, Jeffers suggests that we adopt a cultural constructionism, a view according to which it is (at least partly) cultural differences that make for racial distinctions. In a world without racial injustice, he argues, there could still be people of different races in the sense of people engaged in different kinds of cultural practices. For although race is at least partly a matter of unjust social and political hierarchies, an individual's race is also a central element of their identity, something giving that individual pride and making for bonds with other members of their community:

I think we recognize the value of our shared humanity best when we treat every form of group difference that is meaningful to at least some of its members and which could possibly be benign as an opportunity to prove our ability to benefit from rather than be torn apart by group difference.... Racial identities are meaningful to many of us and are not reducible to positions in a hierarchy. Their preservation should be – and, I believe, is – possible. (Jeffers 2019, p. 201)

Jeffers presents those of us who seek an adequate metaphysics of race with the ethical question: should we want to preserve racial differences in a world without racism? Typically, the social constructionist about race has argued that we need to recognize racial differences today (and not be eliminativists) so that we can identify and correct facts about racism. Jeffers argues that even if there were no racism, we should still care about racial differences. And note this is compatible with the ameliorative project. One can see the cultural constructionism of Jeffers as a descendent of that of W.E.B. Dubois (1868–1963) who argued in his essay “The Conservation of the Races” (1897) that the different races make important contributions to civilization and humanity, in the form of genius, literature, and art, to “help guide the world nearer and nearer that perfection of human life” (1897, p. 78).

There are several challenges we may note for a view like Jeffers's that incorporates a cultural element into a constructionism about race. First, there is the difficulty of making it plausible that there is such a thing as Black, Asian, American Indian, Pacific Islander, and White cultures. Aren't there rather a whole lot of different things which might not have much overlap between them that would count as these cultures? Jeffers

acknowledges that this is a challenge, noting that, for example, not all Black Americans participate in a common culture, narrowly construed. But he thinks it is nonetheless possible to identify some broader cultural elements that tie together all Black Americans. A second challenge that arises for a position like Jeffers's is to make it compelling that these cultures are all ultimately worth preserving. One might wonder, for example, whether there is such a thing we could count as White culture that is worth preserving. White pride is a concept that would have to be wrested away from its supremacist associations to become workable as part of an ameliorative cultural constructionist framework.<sup>6</sup>

Finally, there is a danger of **essentialism** that enters once one introduces a cultural component into one's account of race. By essentialism, we mean the idea that there is a certain essence, or set of properties, that everyone who is a member of a given group must, by necessity, possess. Essentialism is part and parcel of biological realist views about race, but is something that social constructionists have historically wanted to avoid. A member of any one race may share morphologies (hair color, eye color, etc.) and behavioral dispositions with members of any other race. Similarly, members of different races may share interests in art, music, and literature, styles of dress, food traditions. Does the cultural constructionist view risk essentializing, implying that one who does not participate in the culture most often associated with a given race thereby fails to count as a member of that race? This is a question that cultural constructionists like Jeffers will have to address.

**Essentialism:** the view that objects themselves, independently of any way we may categorize them, have certain properties necessarily

#### **EXERCISE 6.4**

##### **Two Forms of Social Constructionism**

This section has introduced two forms of social constructionism about race: socio-political constructionism, which grounds race in a vertical hierarchy of privilege and subordination, and cultural constructionism, which grounds race additionally in features characteristic of different cultures. Which version of social constructionism strikes you as most promising and why?

### **6.5 Gender and Sex**

Some of the difficulties we face when trying to have an adequate metaphysics of race also arise when trying to formulate an adequate metaphysics

of gender. Let's now turn to that topic. First, it will be good to clarify what metaphysicians are interested in when they discuss gender, and how this relates to the metaphysics of sex.

Simone de Beauvoir (1908-1986) famously said, "One is not born, but becomes a woman" (1949, p. 267). One historically significant view aims to follow Beauvoir in the idea that there are two different phenomena at stake when we discuss sex and gender. First, there is sex, a biological phenomenon that one is born into. This includes at least the male sex that is associated with one kind of biology and the female sex, with another. Second, there is gender. To say that one is not born, but becomes a woman is to say that genders are not biological kinds but come about through some kind of socialization process. One's gender status (boy, girl, man, woman, cis, trans, nonbinary, agender) is socially constructed. The sex of a person may influence a person's gender, but does not determine it. Rather, what gender a person falls under is constituted instead by social factors that may include how one self-identifies, how one acts, one's preferences and dispositions, how one is treated and perceived by others, and others' expectations about how one will or ought to act.

If this view, that sex is biological and gender is socially constructed, were the standard view today, then we could just move right on to discuss social constructionism about gender and ask what more specific form this social constructionism should take. However, this is not so. The idea that categories of sex are natural (biological) kinds has been contested by philosophers, including philosophers of science, and biologists. For example, Judith Butler asks:

What is "sex" anyway? Is it natural, anatomical, chromosomal, or hormonal, and how is a feminist critic to assess the scientific discourses which purport to establish such "facts" for us? ... Are the ostensibly natural facts of sex discursively produced by various scientific discourses in the service of other political and social interests? If the immutable character of sex is contested, perhaps this construct called "sex" is as culturally constructed as gender; indeed perhaps it was always already gender, with the consequence that the distinction between sex and gender turns out to be no distinction at all. (Butler 1990, pp. 9-10)

As we will see, some of the arguments that have been raised against biological realism about gender may also be raised to undermine biological realism about sex. So, there is not in the first place agreement that sex is a biological phenomenon. Moreover, as the quotation from Butler indicates, it is not universally agreed that sex and gender are distinct in the first place. There are certainly still philosophers today who think that there is, and many still make use of the sex/gender distinction. But the distinction's value is debated. For the purposes of this chapter, we will focus

on the topic of gender, leaving open the question of whether there is a separate concept of sex, one that is valuable and worthy of metaphysical investigation.

## 6.6 Biological Realism about Gender

As is common when we discuss the metaphysics of social categories, in the case of gender, positions break down into three broad categories: biological realism, eliminativism, and social constructionism. Biological realism is the view that genders are biological kinds (typically, sexes), that there are some biological features that all members of a given gender share in common. Eliminativism is the view that genders are not real, and so classifications of people in terms of genders are the result of a false theory. Social constructionism is the view that genders are real, but that they are social, not natural kinds: their reality is constituted by facts about social practices and institutions. Again, as was the case for race, social constructionism is the dominant position in the metaphysics of gender. But the case against the alternatives to social constructionism about gender looks very different than the case against alternatives to social constructionism about race, and must be considered separately.

Biological realism about gender has historically been equated with **biological determinism**. The thought is that there is something about how men and women are biologically that determines significant facts about their psychologies and behaviors. For example, a biological determinist might believe that men are intelligent, aggressive, and athletic, while women are emotional, submissive, and nurturing, and this is due to facts about their biologies: facts about whether they have or lack certain reproductive organs, the ability (or not) to menstruate and have children, or facts about relative levels of hormones including estrogen and testosterone. There is an analogy here with the assumption of racialism discussed in Section 6.2.

Philosophers as well as biologists have for many decades questioned the idea that there are biological differences that support psychological and behavioral differences between people of different genders. Indeed, Beauvoir's point about one *becoming* a woman was meant to show how psychological and behavioral differences between the genders, and facts about gender inequality, come about predominantly because of social conditioning, rather than innate biological differences. This is an issue that is still widely debated in science as well as in public forums.

The question is then whether biological realism about gender can plausibly be maintained without biological determinism. That is, setting aside the issue of whether biological differences make for interesting differences in intellectual, emotional, or athletic capacities, can one argue that there are nonetheless biological distinctions that map onto gender

**Biological determinism:** the view that the biological features of people of different sexes or genders determines significant facts about their psychologies and behaviors

distinctions? Here, most philosophers are skeptical. There are three kinds of biological differences that are typically appealed to: the possession of XX or XY chromosomes, the presence or absence of specific reproductive organs (testes, ovaries, penis, vulva, etc.), and relative levels of the hormones estrogen and testosterone. There are several problems with biological realisms about gender along these lines. Here, we may focus on two.

First, it is reasonable to think that people generally have an epistemologically privileged access to their own gender. Reflecting on one's own situation, it is difficult to imagine how someone could tell you that your beliefs about your own gender are in error. As Talia Bettcher (2009) argues, people have first-person authority over their gender status. And yet, biological accounts of gender all have the consequence that many people, including most significantly, nonbinary and transgender people, are mistaken about their own genders, because facts about chromosomes, reproductive organs, or hormones fail to map onto the person's justified beliefs about their own gender status. This argument may be stated as follows in the form of a *reductio ad absurdum*:

1. Suppose that biological realism about gender is correct.
2. Then, for each gender, there is a set of biological features that people of that gender instantiate, where their instantiation of those biological features determines that they have that gender.
3. So, for a particular gender G, there will be a set of biological features B that people of gender G instantiate, where the instantiation of B determines that they have gender G.<sup>7</sup>
4. There are some people who possess B but do not believe they are members of gender G.
5. All people have first-person authority over their gender status.
6. Then, it follows from (4) and (5) that there are some people who possess B but are not members of G.
7. It follows from (3) that all people who possess B are members of G.
8. Therefore, biological realism about gender is false.

A biological realist who wants to contest this argument will have to take issue with one of its independent premises, since the conclusion seems to follow validly. These premises are (4) and (5).

### **EXERCISE 6.5**

#### **Gender and First-Person Authority**

How would a biological realist about gender best respond to this argument? What might be said in response?

This objection to biological realism from first-person authority is to many metaphysicians decisive; however, some biological realists may deny that people have privileged access to facts about their own genders, or that one's own beliefs are infallible. So here we may turn to a second objection to biological realism about gender, which turns on the difficulty of finding any biological features that are necessary and sufficient for being a member of a certain gender category. Starting with the possession of XX versus XY chromosomes, biologists have long recognized that some women, those with CAIS (complete androgen insensitivity syndrome), may also possess Y chromosomes. So chromosomal distinctions are not sufficient for gender distinctions. Facts about relative hormone levels are also not sufficient to determine facts about gender, as these are again known to vary among members of different genders, including at birth and as the result of aging. This leaves us with a biological realism based on the presence or absence of certain reproductive organs. We here consider one kind of argument against this version of biological realism, which relies on the fact that human bodies are not neatly divided into two groups each with specific sets of reproductive organs.

Suppose the biological realist identifies men as those with testes and other "male" genitalia, and women as those with ovaries and other "female" genitalia. This raises the question of what to say about the gender status of those humans who are not born with organs matching either of these descriptions. The biologist Anne Fausto-Sterling (1993) has argued that in the human population, we find five kinds of biologies: those with testes, other "male" genitalia and no ovaries or "female" genitalia, those with ovaries and "female" genitalia and no testes or other "male" genitalia, those with both ovaries and testes, as well as other aspects of both "male" and "female" genitalia, those with testes, "female" genitalia, but no ovaries, and those with ovaries, "male" genitalia, but no testes. This would seem to imply that the biological realist should take there to be five genders, not two. However, the philosopher Natalie Stoljar (1995) uses Fausto-Sterling's work to argue against biological realism for the reason that some in the last three groups will be classified (by parents and doctors at birth, or by themselves later) as women, some as men. This implies that facts about which reproductive organs a person has does not provide a consistent biological basis for distinguishing within these categories which are the men and which the women. So facts about which organs a person has do not suffice to determine facts about their gender. This third kind of biological realism also seems to fail.<sup>8</sup>

To this objection, the biological realist might appeal to metaphysical indeterminacy or vagueness (see Section 3.8).<sup>9</sup> Here the claim would be that those with testes, other "male" genitalia and no ovaries or "female" genitalia are men, those with ovaries and "female" genitalia and no testes or other "male" genitalia are women, and those with other combinations



of reproductive organs have genders that are metaphysically indeterminate. This is a coherent position, however Stoljar and other social constructionists about gender will here reiterate that most of the people this account would claim to be of indeterminate gender are taken to have completely determinate genders by their doctors, families, and themselves.

## 6.7 Essentialism or Eliminativism?

But if genders are not to be understood in terms of the sharing of some common biological features, then what is it that makes for the sharing of a common gender? In the twentieth century, feminist theorists tried to answer this question, focusing especially on what it is that ties together the class of women. One aim was to propose an account of what it is to be a woman that would respond to accounts of biological determinists according to which women are, to use Beauvoir's terminology "the second sex," lesser both intellectually and morally.<sup>10</sup> Feminist theorists argued that women weren't second to men in these respects, and made proposals involving what one might arguably take to be the special characteristics of women. For example, to cite two very different feminist accounts, in *The Feminist Mystique*, Betty Friedan (1921-2006) critically discussed a conception of women as tied to family life and the home, and Adrienne Rich (1929-2012) proposed understanding the female essence in terms of "our great mental capacities ... our highly developed tactile sense; our genius for close observation; our complicated, pain-enduring, multi-pleasured physicality" (1977, p. 290).

These proposals were attractive and liberating for some women at the time. However, they face difficulties when taken as answers to the metaphysical question of what it is to be a woman ("the woman question" in feminist theory). Not all women have or want families, not all are especially tactile or possess an attention to detail. There doesn't seem to be any one common set of intrinsic features that all women everywhere possess. In the 1980s, feminist theorists thus turned away from proposals of this kind, those seeking to identify intrinsic features that would make someone a woman or characterize a woman's essence, not least as it was pointed out that the various proposals for how to understand the nature of and distinctive problems faced by women were largely being formulated by a particularly undiverse group of women: white, upper-middle class academics, claiming to speak for all women (hooks 1982, Spelman 1988). In the 1980s, this was termed the **problem of essentialism**. As Linda Martín Alcoff has noted, it seems that any way of proposing an account of what it is to be a woman, biological or not, faces this problem:

To the extent it reinforces essentialist explanations of these attributes, it is in danger of solidifying an important bulwark for sexist

### Problem of

**essentialism:** a problem for accounts of gender of reinforcing essentialism, so that those who do not meet the conditions of the account are lesser or inferior members of that gender

oppression: the belief in an innate “womanhood” to which we must all adhere lest we be deemed either inferior or not “true” women. (Alcoff 2006, p. 139)

In proposing a metaphysical account of a social category, especially a human category, we risk imposing essentialist conditions on what it is to be a member of that category that many do not in fact possess. It is difficult to imagine there are any intrinsic features we can claim that all women have.

In response to this concern, Stoljar proposed a “cluster concept” view of gender. The idea of a **cluster concept** comes from the work of Wittgenstein (1958). Wittgenstein noted that for many categories, although there is no set of necessary and sufficient conditions that all things that fall under the category must possess, the things that fall under that category do bear a family resemblance to one another. That is to say, there is a cluster or set of features characteristic of members of that category such that all members possess at least some of those features.

Stoljar (1995) proposes that one is a woman if one possesses enough of the following four kinds of characteristics:

1. biological and bodily features associated with the human female sex (XX chromosomes, female genitalia, female voice quality or gait);
2. phenomenological features, i.e. aspects of what it “feels like” to be a woman, including physical feelings (e.g., the pain of menstrual cramps or giving birth) and emotions (e.g., fear of walking on the streets at night);
3. features of occupying kinds of social roles including wearing certain kinds of clothes, having certain kinds of occupation, being oppressed on the basis of one’s sex;
4. attributions of womanness, both self-attributions and attribution by others.

As is standard with cluster concept accounts, these kinds of characteristics are then taken to determine what makes for a paradigm woman. According to Stoljar, “Any individual resembling any of the paradigms sufficiently closely . . . will be a member of the resemblance class woman” (1995, p. 284).

Although this account succeeds in capturing the fact that there seems to be no set of intrinsic features (biological or otherwise) that are common to all women, many metaphysicians of gender find it to fail for another reason, namely by relying on the assumption that there is such a thing as a paradigm of woman. One might worry in the first place that many of the criteria making for one’s instantiation of the paradigm of womanness rely on misogynist ideals, that women dress or act in a certain way, that they take on certain occupations (of caregiving) and not

**Cluster concept view:**

a view about certain categories that although there is no set of necessary and sufficient conditions that all things that fall under the category must possess, the things that fall under that category do bear a family resemblance to one another, in possessing enough of a set of features

others. But additionally, as Bettcher (2011) argues, such a cluster concept view has the consequence of marginalizing those women who meet the criteria to a lesser degree, not counting them as women. Thus, shifting from an essentialist view of woman to one according to which woman is a cluster concept does not adequately respond to Alcoff's worry about essentialism.

It is at this point that eliminativism looms. If there are no features we can point to that are shared by all women (*mutatis mutandis* for men and other genders), and a view according to which women are those who share all of a cluster of certain features seems also to be mistaken, then perhaps we might do better to reject that there is anything for our gender concepts to refer to. And so, we ought to reject these categories altogether.

The trouble, however, is that there are costs one incurs by adopting eliminativism about gender. These costs are similar to those that are incurred by those who endorse eliminativism about race. First, individuals' genders are causally potent, they have real effects on the kind of lives people live, both for better, in terms of the roles they play in constituting our personal identities and the opportunities they afford us (for example, for connection and bonding with others), and also for worse, in terms of the discrimination, oppression, and constraints those of especially marginalized genders face. If we accept the Eleatic principle that says what is causally potent is real, then this implies genders are real.

Additionally, it is central to the feminist project that we be able to identify gender-based discrimination and oppression where it arises in the first place, in order to overcome it. However, if there are no genders, if these are empty concepts, how can there be gender-based discrimination and oppression?

With this in mind, Alcoff has argued that those wanting a metaphysical account of gender seem to face a dilemma: adopting intrinsic characterizations of gender categories leads one to an essentialism that fails to count all women as women, but rejecting the categories altogether makes one unable to identify and address gender-based injustice.<sup>11</sup> The solution Alcoff proposes is that we adopt a **positional view about gender**. Genders are real, but having a gender isn't having a set of intrinsic features, it is rather occupying a kind of social position. This implies a relational, rather than an intrinsic characterization of gender:

The concept of positionality allows for a determinate though fluid identity of woman that does not fall into essentialism: woman is a position from which a feminist politics can emerge. Seen in this way, being a "woman" is to take up a position within a moving historical context and to be able to choose what we make of this position and how we alter this context. (Alcoff 2006, p. 149)

**Positional view of gender:** the view that genders are real, but having a gender isn't having a set of intrinsic features, but rather occupying a kind of social position

This has been an influential starting point and we will see in the next section how it is adopted by several others in the debate over the correct metaphysics of gender.

## 6.8 Social Constructionism about Gender

One metaphysical account that makes one's gender a social position is that offered by Haslanger. We can see in the following that Haslanger's proposal about gender has many of the same features as her proposal about race:

S is a woman if

- (i) S is regularly and for the most part observed or imagined to have certain bodily features presumed to be evidence of a female's biological role in reproduction
- (ii) that S has these features marks S within the dominant ideology of S's society as someone who ought to occupy certain kinds of social position that are in fact subordinate (and so motivates and justifies S's occupying such a position); and
- (iii) the fact that S satisfies (i) and (ii) plays a role in S's systematic subordination, that is, *along some dimension*, S's social position is oppressive, and S's satisfying (i) and (ii) plays a role in that dimension of subordination. (2012, p. 234)

Haslanger proposes an analogous account of what it is to be a man, but let's focus on this account of what it is to be a woman. We see the positional component explicitly in clauses (ii) and (iii), with S's perceived or imagined bodily features and facts about the dominant ideology of her society causing her to occupy a kind of position, one in which she is subordinated along a certain dimension. This is an account that is also tailored to accommodate **intersectionality**, by recognizing that although a person may be subordinated along one dimension (by virtue of being perceived or imagined to have certain reproductive organs), one may be also subordinated or privileged along others, in virtue of one's status with respect to race, disability, age, and so on. Note as well that although there is a biological component to this definition, a person's satisfaction of clause (i) does not require that they actually possess any specific set of organs. It suffices for others to imagine them to possess these organs.

Haslanger appears at first to avoid Alcoff's dilemma. The account does not introduce any list of intrinsic features (biological, psychological, or otherwise) one must possess in order to be a woman. As such, it does not appear to face the charge of essentialism. Nonetheless, one may be concerned that it does face a mismatch problem (see Section 6.2). For is it the case that all women are such that they are for the most part

**Intersectionality:** the fact that the same person may be a member of several different social categories, so that they may be subordinated along several different dimensions, or subordinated in some respects, but privileged in others

observed or imagined to have certain bodily features presumed to be evidence of a female's biological role in reproduction, and in virtue of this fact, oppressed along a certain dimension? This seems not to be so.

There are two sorts of case one might worry about. First, there is the Queen of England case (see Mikkola 2017). This is the problem for Haslanger's account that there are women that are not in fact oppressed at all, but rather have a great deal of privilege, indeed occupying a position in society higher than any men. Haslanger's account will fail to count them as women. And then there are other women, for example some trans women, who do not publicly present themselves as women or who do but whose gender presentation is not respected, who are oppressed, but not because they are for the most part observed or imagined to have certain bodily features presumed to be evidence of a female's biological role in reproduction. Again, Haslanger's account will not count these women as women (Jenkins 2016). Moreover, Haslanger's account seems to count some trans men and nonbinary people as women, those that are systematically oppressed on the basis of being for the most part observed or imagined to have certain bodily features presumed to be evidence of a female's biological role in reproduction. Note that this relates to a general challenge for positional accounts. It is questionable whether there is a common social position that may be identified which all women in a given society, even at a given time, occupy.

The different kinds of problem just stated require different kinds of treatment. Let's consider the Queen of England case first. To respond, Haslanger may note that her account is proposed in the context of an ameliorative approach to the question of "What is gender?" Haslanger is aiming to develop an account of gender that can fulfill feminist and anti-racist aims of identifying and ending gender-based injustices. Recall from Chapter 5 that accounts aimed at amelioration will generally be revisionary. As such it might not count every person we would ordinarily count as a woman as a woman (and it may count some we might not ordinarily count as women as women). This is fine, assuming the account achieves its political aims. Given this fact, Haslanger can argue that the Queen of England case is not a problem for her approach. What is important for having a metaphysical account that can address gender injustice is having an account of women that includes those women who are the subject of gender-based oppression. One might argue it is not necessary to include those who are not subject to such oppression as women.

As Katherine Jenkins has argued, the same response cannot justify leaving trans women out of the scope of the account, for they are the subjects of gender-based oppression and as such, an ameliorative approach to gender should include them as women. Jenkins's proposal is that we move to an account of gender based in facts about gender identities: the

women are those who have a female gender identity. Different metaphysicians have proposed different accounts of what it is to have a female gender identity. Haslanger herself has proposed we think of gender identities in terms of internalized norms of what is appropriate gendered behavior (2012, p. 228). Jenkins proposes the following account:

S has a female gender identity iff S's internal "map" is formed to guide someone classed as a woman through the social or material realities that are, in that context, characteristic of women as a class. (2016, p. 410)

There are several components to this definition that need unpacking.

First, when Jenkins discusses being "classed as a woman," she means one is such that the dominant ideology of their society counts them as a woman. She then considers a person's gender identity in terms of an internal or mental "map," one that may be conscious or may largely be unconscious, that guides one as one navigates their environment. This environment is material, in being made up of physical objects one interacts with, but it's also social: one interacts with other people who have certain expectations about how people classed by the dominant ideology as women will behave, about how they should behave. Importantly, one can have a female gender identity even if one is not classed by the dominant ideology as a woman.

Jenkins introduces a helpful example to illustrate her notion of gender identities as internal maps. It is part of the dominant ideology in the U.S. that those who are classed as women are expected to remove the hair from their legs, whereas those who are classed as men are expected to keep the hair on their legs. One who has a female identity will then be such as to have a way of responding to this norm that differs from one with a male identity. This way of responding to the norm may be by shaving, but it also may be by flouting the norm and choosing not to shave her legs. As Jenkins puts it, "her experience of social and material reality includes navigating the norm that women should have hairless legs, even though she is not complying with it" (2016, p. 411).

Ultimately, Jenkins argues, it is not essential to the ameliorative project that we completely reject other accounts of gender, such as Haslanger's. We may recognize that there is a category of people who are oppressed based on the reproductive organs they are observed or imagined to have. However, it is important as well to recognize that there are also other categories of people who face gender-based injustices, one such group are those who have female gender identities. And so ameliorative accounts should recognize this category as well. Jenkins argues thus that the best ameliorative account of gender will be a pluralistic one that recognizes both the category of women as picked out by Haslanger's account and the

category of those who have female gender identities. And we should have an account that also makes room for those who have nonbinary, fluid, or mixed gender identities. It is not clear why a metaphysics of gender must be monistic.

## 6.9 Conferralism about Gender

We can discuss one other way of responding to the mismatch problems many have raised for Haslanger's account. This is by adopting the conferralist framework proposed by Ásta in her book *Categories We Live By*. According to Ásta, one has a particular gender status if subjects with standing in a particular context perceive that one has certain properties (the "base properties") relevant to being a member of that gender category. What base properties are relevant depends on the context and what the members of the context with social standing take to be relevant to gender status. To be a member of a given gender, one need not actually possess the relevant base properties; it is just essential that others perceive one to possess those features. This is a conferralist account in that genders are social statuses that are conferred onto people by others with standing in a context.

The conferralist account of gender is similar to Haslanger's in the respect that one's having a gender status is a social feature of that person, one they have in virtue of how they are perceived by other people. But there are two features of Ásta's framework that make it different from Haslanger's and able to avoid the mismatch objections. First, the account is not hierarchical. Ásta gives the following examples of what the base properties relevant to gender status may be: role in biological reproduction, sexual engagement, bodily presentation, self-identification. In conferring the social status of being (e.g.) a woman on someone, one is not necessarily conferring a place in a social, political, or economic hierarchy. Thus, Ásta's account can allow that in some contexts, the Queen of England is a woman, for example those in which the relevant base properties are bodily presentation or self-identification. Moreover, the account has it that trans women are women, even those who do not present themselves as women or whose presentations are not accepted, since there are contexts in which the gender status woman is conferred on trans women, in virtue of their possessing other relevant base properties.

Although the conferralist framework has the ability to handle some mismatch problems, there are other problems that seem to arise for it. Although it has the consequence that there are contexts in which trans women are women, since there exist contexts in which people with standing confer the gender status of women on trans women, the account also has the consequence that there are contexts in which some trans women

are men, since there are contexts in which people with standing confer the gender status of men on trans women. Since the account is (intentionally) radically contextual, there is no way to say, on the conferralist account that one of these conferrals is right, that it gets the metaphysics of gender right, and the other is wrong.

Ásta herself accepts this consequence of her account. She argues that the metaphysics of gender does not run deeper than our conferrals and so there are no real facts about gender to settle which conferrals are (descriptively) right and which (descriptively) wrong (2018, p. 90). Ásta acknowledges that there is a normative question we may ask. This is whether gender categorizations that count trans women as men are morally or politically justified. And if we answer “no” to this question, then we should be working to move towards a society in which those with social standing don’t confer this status on trans women. As Ásta puts it, “on the account I offer, the question *what is a woman?* is thus revealed to ultimately demand an answer to a different question: *who should be a woman?*” (2018, p. 90).<sup>12</sup> But, in her view, these moral and political questions are separate from the metaphysical question her conferralist account aims to address.

## **6.10 A World Without Gender?**

Finally, just as in the case of race, we can ask in the case of gender what a metaphysical account should imply for the situation in which feminism has achieved its goals and there is no more gender-based injustice.

Accounts like Haslanger’s that take women to be those who are oppressed on the basis of their perceived or imagined reproductive status will imply that in a world like that, there will be no more genders, and so no more men and no more women. This may be a welcome result for some, who see their gender merely as a barrier to equal treatment. However, we must recognize as well, that just as Jeffers points out in the case of race, one’s gender is often an integral and valued part of one’s identity. Moreover, trans and nonbinary people have had to fight, indeed continue to fight, to have their gender status recognized. Given this fact, it is hard to see how a world of justice is a world where no one has a gender status.

This point seems also to threaten Jenkins’s gender-as-identity account since on this account, gender identities depend on there being objective facts about how the dominant ideology treats those it classes as men and women. If there is no dominant ideology that treats men and women differently, then there is no anchor for what these gender identities are. This raises the question of whether an alternative identity-based account can be developed, or whether this is an inevitable consequence of social constructionist views.

In any event, metaphysicians of gender (and race) have been making much progress developing multiple accounts of gender (and race) that can



best serve us in the world in which we live, one in which we have not yet reached an end to gender and racial injustice. It may be difficult to wrap our minds around what would be best in a world where such justice is eradicated. But we must remember that insofar as these metaphysicians are adopting the ameliorative approach, their first goal will be to develop accounts that can serve us now in our current circumstances.

### EXERCISE 6.6

#### Beyond Gender Categories

In your view, in a world without gender-based injustice, what would make for the distinction between different genders? Or in such a world, would eliminativism about gender be the correct position?

### Suggestions for Further Reading

An excellent recent source on the metaphysics of race is *What is Race? Four Philosophical Views*, by Joshua Glasgow, Sally Haslanger, Chike Jeffers, and Quayshawn Spencer. For defenses of eliminativism, see Kwame Anthony Appiah's *In My Father's House: Africa in the Philosophy of Culture* and Joshua Glasgow's *A Theory of Race*. For more on the science of race, see Naomi Zack, *Philosophy of Science and Race*.

On the metaphysics of gender, see the article on "Feminist Approaches to Sex and Gender" by Mari Mikkola in the *Stanford Online Encyclopedia*. See also the essays on both race and gender in Linda Martín Alcoff's *Visible Identities* and Sally Haslanger's *Resisting Reality*. Ásta's *Categories We Live By* and Charlotte Witt's *The Metaphysics of Gender* are book-length treatments of the topic.

### Notes

- 1 One may note that Hispanic is often recognized to be an ethnicity, not a race, and that people who identify as Hispanic will often classify themselves as one or more of the races in Spencer's biological theory (White, Black, and /or American Indian). However, data cited by Glasgow (OMB 2016) indicates that nearly half of those who identify as Hispanic would not identify as members of one of these races. Reporting after the 2020 U.S. census shows that "Other" became the second highest racial category respondents identified as ([www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html](https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html)). Some have interpreted this as supporting Glasgow's conclusion that those identifying as Hispanic do not identify as one of the five races in Spencer's theory, but as a distinct race not recognized as one of the five human continental populations. See, for example, [www.npr.org/2021/09/30/1037352177/2020-census-results-by-race-some-other-latino-ethnicity-hispanic](https://www.npr.org/2021/09/30/1037352177/2020-census-results-by-race-some-other-latino-ethnicity-hispanic).

- 2 Not all biological realists agree. Andreasen has argued "the existence of biological races does not depend on our folk taxonomy being right" (1998, p. 213).
- 3 For a recent account, see Wilkerson (2020). Though see Mallon (2016, Chapter 1) for an argument based on findings from cognitive science that racial classification is instead innate in all humans.
- 4 The sense of 'witch' under discussion here is different from the sense in which some identify as Wiccan today.
- 5 Wilkerson (2020).
- 6 Here Jeffers notes there may be a way to do so. See Linda Alcoff (2015).
- 7 We may insert here for 'G' any gender that the biological realist countenances as real.
- 8 Ásta (2018), pp. 71-72, uses this same biological work to develop an argument against biological realism about sex, arguing that it too is socially constructed. As she puts it, "the determination of which physical properties are important for sex assignment, and in particular, the assignment of people into one of two sexes, is shaped by societal values and interests" (2018, p. 72).
- 9 Another response the biological realist might make to this argument is to shift to a cluster concept view, saying that to be a man (or woman or member of another gender), one only needs to have enough of the biological features characteristic of men (or women or members of the other gender). The cluster concept view is described and evaluated in Section 6.7.
- 10 Alcoff (2006), Chapter 5 provides a useful overview of these proposals.
- 11 See also Bach (2012).
- 12 See also Dembroff (2018). Dembroff argues that the important questions are not what are the metaphysically real facts about who counts as what gender, but what are the best classification practices to use. The gender kinds operative in dominant contexts may themselves be oppressive. But in these cases, we shouldn't deny their reality, but work to change them.

# 7 Grounding and Fundamentality

## Learning Points

- Presents the mind-body problem as a central example of a fundamentality question
- Introduces supervenience and its application to the formulation of philosophical positions
- Introduces Schaffer's grounding framework and some standard challenges to it
- Compares grounding with other notions of ontological dependence
- Discusses two ways to understand what it means to say that some entities are fundamental
- Considers the question of whether there is a fundamental level to reality

## 7.1 Fundamentality Questions

In this chapter, we move beyond ontological questions - questions about what kinds of entities do or do not exist - to consider a different set of metaphysical questions. These are questions concerning which (if any) among the entities that exist are fundamental, and what are the relations that obtain between the fundamental and nonfundamental. Metaphysicians will ask, for example: are physical entities like electrons and atoms more fundamental than minds and other mental phenomena? Are physical entities more fundamental than numbers and other mathematical entities? What about the laws of nature? Are the laws of nature ultimately fundamental entities or do they themselves depend on physical entities like electrons and atoms? What about moral and aesthetic properties of individuals and

actions? Are they ultimately fundamental, or are they somehow less fundamental than physical properties, or properties of minds?

In this chapter, we introduce several technical concepts that metaphysicians have developed in order to frame discussions of fundamentality. We will start by investigating those used to formulate and debate the answers to questions about **relative fundamentality**: questions concerning how or whether one kind of entity might be more fundamental than another kind. The mind-body problem is a central example of this kind of metaphysical debate, and so we will use this problem to illustrate the key concepts of relative fundamentality, which include supervenience, realization, and grounding.

**Relative fundamentality:** one entity or fact's being more, less, or equally fundamental than some other entity or fact

## 7.2 The Mind-Body Problem

Traditionally, the mind-body problem is the question of how one entity, one's mind, is related to another, one's body. In answering this question, some follow the seventeenth century natural philosopher Rene Descartes and adopt **interactive substance dualism**, a view according to which mind and body are made of two numerically distinct and equally fundamental kinds of substance (mental substance and matter, or material substance) that causally interact with each other.<sup>1</sup> Others adopt a metaphysics according to which one substance is more fundamental than the other. For example, **materialism** is the view that matter is fundamental. Materialists traditionally think that minds are nonfundamental and somehow built up out of more fundamental, material entities such as brains or bodies. Finally, **idealism**, an extremely influential view over the history of philosophy from Plato through the nineteenth century, has it that minds are fundamental. If there are material bodies, then they are nonfundamental entities somehow constructed out of mental ones (ideas).<sup>2</sup>

**Interactive substance dualism:** the view that minds and bodies are distinct kinds of substances that causally interact with one another

**Materialism:** a historical predecessor to contemporary physicalism, the view that matter alone is fundamental

**Idealism:** the view that minds or mental entities (ideas) alone are fundamental

Note that the main question of the mind-body problem is not existential: it is not about whether minds and bodies exist. Traditionally, the substance dualist, materialist, and idealist all take it for granted that we have minds and bodies (or perhaps, we *are* minds and/or bodies). The question, rather than being whether minds and bodies (or mental and material/physical substances) exist, is which if any are more fundamental and what is the relation between the two.

Today, when philosophers debate the mind-body problem, most of the time, they do not frame it like Descartes did as a question about the relation between two kinds of substance: mental and material. There are some who do, especially contemporary substance dualists. But many doubt that the best way to think about minds is in terms of a special

**Category mistake:** a mistake that involves placing some entity in the wrong ontological category

kind of substance. Following Gilbert Ryle (1900-1976) in his 1949 book *The Concept of Mind*, many today believe any claim that there is such a thing as a mental substance makes what philosophers call a **category mistake**. It is a mistake of placing the mind or mentality into the wrong ontological category. It can be debated which ontological category is more appropriate for formulating the mind-body problem, but many today frame it in terms of the category of properties. So, the central question is: what is the relationship between mental properties (like being self-aware, having certain beliefs or desires, or having certain sensory experiences) and physical properties (such as neurophysiological properties or other properties of one's body)?

### EXERCISE 7.1

#### Category Mistakes

Ryle argued that it was a category mistake to view minds as substances or things. He argued that having a mind wasn't the possession of a thing, like owning a house or a pair of pants. Rather having a mind was more like having the capacity for a certain kind of behavior. Minds are more like dispositions than substances or things.

What are some other phenomena, beyond mental phenomena, that a metaphysician might argue are mistakenly put in the category of substances or things?

### TEXT BOX 7.1

#### Ontological Categories

The idea that different entities may belong to distinct ontological categories was noted already in the fourth century BC by Aristotle (384 BC-322 BC). In Aristotle's *Categories*, he claims that there are ten different ontological categories. These are: substance, quantity, quality, relatives, place, time, position, way of having, way of acting, and way of being acted upon. Aristotle illustrates these categories in the following way:

To give a rough idea, examples of substance are man, horse; of quantity: four-foot, five-foot; of qualification: white, grammatical; of a relative: double, half, larger; of where: in the Lyceum, in the market-place; of when: yesterday, last-year; of being in a position: is-lying; is-sitting; of having: has-shoes-on, has-armor-on; of doing: cutting, burning; of being affected: being-cut, being burned. (1b25)

Today, many metaphysicians might adopt a different set of ontological categories. For example, many would think of some of Aristotle's categories (quantities, qualities, position, and having, at least) as different species of the same category of property. Some would also defend the existence of additional categories motivated by mathematical and scientific developments that have occurred since the fourth century BC. For example, one might argue that rather than seeing place and time as distinct categories, physics motivates us to think in terms of a new category of spatiotemporal location. The same physics, it has been argued, motivates us also not to think in terms of the ontological categories of substance and property, but instead event. This was an argument Bertrand Russell developed in his 1927 book *Analysis of Matter*, and was followed up on by W.V. Quine (1950) and the perdurantist philosophers we will discuss in Chapter 9.

Some metaphysicians would disagree that science may force us to revise our beliefs about which ontological categories there are. E.J. Lowe (1950-2014), for example, has argued that the question of which ontological categories there are is a matter to be settled independently of science, that is, a priori. Lowe himself argues for a four-category ontology of substantial particulars, non-substantial particulars, substantial universals, and non-substantial universals. Others have thought a priori reasoning leads us to a more narrow range of categories. For example, trope theorists (see Chapter 2) like D.C. Williams (1899-1983) and those who follow him think we can understand everything there is in terms of one fundamental category of tropes or abstract particulars (Williams 1953). Today, there are a range of views about what the correct division of ontological categories are in nature. However, one might still think that Aristotle's list at least gives us a reasonable starting point.

Here again, there are three main views. Mental properties and physical properties are numerically distinct and equally fundamental (**property dualism**). Physical properties are more fundamental than mental properties (physicalism). Mental properties are more fundamental than physical properties (**phenomenalism**). Physicalism tends to be the most popular position, although many philosophers also subscribe to property dualism. Fewer philosophers tend to be phenomenologists, although there has been a renewal of interest in the view recently.<sup>3</sup> In what follows, we will primarily focus on different versions of physicalism. This will allow us to illustrate the different senses in which one set of properties might be more fundamental than another set. Physicalist philosophers continue to debate the correct way to make sense of the relative fundamentality of physical properties vis-à-vis mental properties.

Initially, in the 1950s and 1960s, the most common version of physicalism was reductive physicalism, also known as type physicalism, and also

**Property dualism:**

the view that mental properties and physical properties are numerically distinct and equally fundamental

**Phenomenalism:**

the view that mental properties are more fundamental than physical properties

**Identity theory of mind:** the view that mental properties are numerically identical with a certain type of physical properties, neurophysiological properties

**Nonreductive physicalism:** the view that although physical properties are more fundamental than mental properties, mental properties are numerically distinct from physical properties

known as **the identity theory of mind**. This is the view that mental properties are numerically identical with a certain type of physical properties, neurophysiological properties.<sup>4</sup> So, for example, to say that I am having a particular type of sensory experience is just to say that I am instantiating a particular type of brain state.

Text Box 7.2 describes the central argument that drove many philosophers away from the identity theory and toward what is called **nonreductive physicalism**. Nonreductive physicalists believe that mental properties are numerically distinct from physical properties. We can't identify the having of any particular type of sensory experience with one's being in any particular type of brain state. But still, the nonreductive physicalist wants to say, physical properties are more fundamental than mental properties. One couldn't instantiate a mental property without instantiating some or other underlying physical property. The question is then, what is the metaphysical relationship between mental and physical properties, if it is not numerical identity.

#### TEXT BOX 7.2

##### The Multiple Realization Argument

The identity theory of mind, also known as type physicalism, is the view that mental properties are identical with physical properties. According to the identity theory, for example, *being in pain* is not a sui generis mental property, but rather is identical with some particular physical property, one most likely to be identified by neuroscience. The identity theory was a view advocated in the 1950s by U.T. Place (1924–2000), Herbert Feigl (1902–1988), and J.J.C. Smart (1920–2012). Probably the most famous argument for the view, presented by both Feigl (1958) and Smart (1959), appealed to Ockham's Razor: According to Smart and Feigl, we should believe the identity theory because it is the most ontologically parsimonious way to account for the pervasive mental-physical correlations discovered by neuroscience. The most parsimonious explanation for any persistent correlation we find between a mental property M and a physical property P is that M is identical with P. So, we should not believe that M and P are two numerically distinct properties.<sup>5</sup>

Although the identity theory is an appealingly simple view of the mind-body relation, in the next decades, philosophers of mind developed arguments that appeared to undermine the view. The

most famous and influential such argument, presented in the 1960s and 1970s in work by Hilary Putnam (1926–2016) and then Jerry Fodor (1935–2017), was the **multiple realization argument**. This is an argument against the identity theory (type physicalism) and in favor of some form of nonreductive physicalism, especially realization physicalism.

The multiple realization argument is pretty simple. Suppose for reductio that the property of being in pain *is* identical to some specific neurophysiological property P. Then this would imply that any creature that is in pain must instantiate that same neurophysiological property. But, Putnam (1967a) and Fodor (1974) argued, it is plausible that psychological properties like being in pain are instantiated by creatures with a wide variety of anatomies. For example, it seems plausible that not only humans and other mammals, but also birds, reptiles, even cephalopods like octopi can be in pain. So maybe humans that are in pain instantiate property P, but given their different anatomies, it seems more plausible that octopi that are in pain instantiate some very different neurophysiological property; call it 'P\*.' If so, this undermines the identity theory, since it means that being in pain cannot be identical to any specific neurophysiological property. Instead, we should believe, if we are physicalists, that all instances of pain are realized by some or other physical property, but not necessarily by the same physical property in all creatures. As Fodor puts it:

There are no firm data for any but the grossest correspondence between types of psychological states and types of neurological states, and it is entirely possible that the nervous system of higher organisms characteristically achieves a given psychological end by a wide variety of neurological means. If so, then the attempt to pair neurological structures with psychological functions is foredoomed. (1974, p. 105)

Thus, psychological or mental properties are multiply realizable in physical properties, and thus are not numerically identical to physical properties.

So, many physicalists argue that the correct way to view the relationship between mental and physical properties is not as one of identity, but rather realization. But what does it mean to say mental properties are **realized by** physical properties? What is realization? Here, we will just present one standard understanding

**Multiple realization argument:** an argument against type physicalism stating that because the same mental property can be and often is realized by distinct physical properties, mental properties cannot be identical to physical properties

**Realization:** a relation between two properties A and B such that B is the property that plays the causal/functional role characteristic of property A



of realization by Jaegwon Kim (1998).<sup>6</sup> For Kim, a property M is realized by a property P just in case:

- (1) The property M can be given a functional characterization, that is, a characterization in terms of its causal/functional role. This is to say, something x instantiates property M just in case x instantiates some property or other that plays a particular causal role R; and
- (2) There is some property P that plays role R when it is instantiated by x.

A causal role may be characterized in terms of a specification of inputs and outputs. For example, if M is the property *being in pain*, then the causal role R might be being a property whose instantiation is typically caused by tissue damage and typically results in winces and groans. Neuroscientists may then be able to discover the physical property P whose instantiation is caused by tissue damage and results in an organism wincing and groaning. If pain is multiple realized, this means that distinct properties (P and P\*) may play this causal role in different kinds of creatures.

### 7.3 Supervenience

Starting in the 1970s, the most common way for physicalists to explain what it could mean that mental properties ontologically depend on physical properties without being strictly numerically identical to them was by adopting a supervenience thesis. Strictly speaking, the claim that mental properties supervene on physical properties is compatible with the identity theory, since if mental properties are identical with physical properties, then, as we will see, they trivially supervene on physical properties. But often, when a philosopher endorses a supervenience claim between two sets of properties, they are meaning to emphasize that the relation between these properties is weaker than identity.

So, what is it to say that some properties supervene on some other properties? Text Box 7.3 presents the official formal definitions of **supervenience** with which metaphysicians tend to work, but it is also helpful to consider three less formal glosses on the notion.

**Supervenience:** a relation between two sets of properties A and B such that the instantiation of the A-properties completely fixes the instantiation of the B-properties

1. The B-properties supervene on the A-properties just in case the instantiation of A-properties completely fixes the instantiation of B-properties.
2. The B-properties supervene on the A-properties just in case there can be no change in the instantiation of B-properties without a change in the instantiation of A-properties.

3. The B-properties supervene on the A-properties just in case entities that share all of their A-properties are also duplicates with respect to the B-properties they possess.

Substituting 'mental properties' for 'the B-properties' and 'physical properties' for 'the A-properties,' we get:

(1MP) Mental properties supervene on physical properties just in case the instantiation of physical properties completely fixes the instantiation of mental properties.

(2MP) Mental properties supervene on physical properties just in case there can be no change in the instantiation of mental properties without a change in the instantiation of physical properties.

(3MP) Mental properties supervene on physical properties just in case things that are physical duplicates are also mental duplicates.

Note that generally, supervenience physicalists intend their thesis to be about more than just mental properties. Supervenience physicalists tend to think that all (or all contingent) properties instantiated in our world (for example, the colors and other observable features of material objects, the aesthetic properties of works of art or architecture, the ethical characteristics of actions) supervene on the physical properties.<sup>7</sup>

The supervenience physicalist believes that once one fixes the physical properties, one fixes all of the properties that are instantiated at our world. And, there can be no change in the pattern of instantiation of any of these properties (and so anything that happens) without there being some change in the pattern of instantiation of physical properties. Some supervenience physicalists put things this way: all God had to do when He created the world was to fix the instantiation of the physical properties. This is because if all of the other properties supervene on the instantiation of the physical properties, then once God fixes the pattern of instantiation of the physical properties, the instantiation of all of the other properties is determined; they come along "for free." Supervenience physicalists will sometimes use the metaphor of the non-fundamental, supervenient features of the world being an "ontological free lunch."<sup>8</sup>

### **TEXT BOX 7.3**

#### **Concepts of Supervenience**

In his 1984 paper "Concepts of Supervenience," Jaegwon Kim distinguishes between what he calls weak and strong forms of

**Weak Supervenience:** A weakly supervenes on B if and only if necessarily, for any x and y, if x and y share all properties in B, then x and y share all properties in A

**Strong Supervenience:** A strongly supervenes on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and *necessarily* if any y has G, it has F

**Global Supervenience:** The A-properties globally supervene on the B-properties just in case worlds that are indiscernible with respect to the instantiation of B-properties ("B-indiscernible," for short) are also indiscernible with respect to the instantiation of A-properties (A-indiscernible)

supervenience. Where A and B denote families of properties (e.g., the mental and the physical properties), he defines weak supervenience in the following way:

**Weak Supervenience:** A weakly supervenes on B if and only if necessarily, for any x and y, if x and y share all properties in B, then x and y share all properties in A – that is, indiscernibility with respect to B entails indiscernibility in A.

Kim argues that weak supervenience is too weak to capture physicalism. This is because although it ensures that all actual physical duplicates are mental duplicates, it is compatible with the possibility of other creatures with the same physical properties having very different mental lives, or possessing no mental properties at all. For this reason, he proposes physicalists adopt mind-body supervenience in a stronger form:

**Strong Supervenience:** A strongly supervenes on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and *necessarily* if any y has G, it has F.

Mind-body supervenience in this sense ensures that any possible creatures whatsoever that share the same physical features also share the same mental features.

Kim notes that these definitions of supervenience make supervenience a *local* matter. It is local in the sense that these definitions refer to the properties entities or individuals (x or y) instantiate, rather than the properties instantiated at entire worlds or universes. But one might wish to instead adopt a more global supervenience claim of this form:

**Global Supervenience:** The A-properties globally supervene on the B-properties just in case worlds that are indiscernible with respect to the instantiation of B-properties ("B-indiscernible," for short) are also indiscernible with respect to the instantiation of A-properties (A-indiscernible).

In his 1984 paper, Kim argued that weak supervenience was too weak to state physicalism. He believed the strong and global varieties were better able to characterize the position, although ultimately, as we will see, Kim thought that additional notions of ontological dependence were needed.

We can illustrate what is entailed when one endorses a supervenience thesis with the following example. Suppose one wanted to argue that the aesthetic properties of objects all supervene on their physical features. What would one be committed to? If one adopted a **local supervenience** thesis about the aesthetic, then this would mean that once one fixes the physical properties of a work of art such as a painting, all facts concerning whether the painting is beautiful or not are fixed. And this would mean that any two paintings with the same physical characteristics would be equally beautiful. There could be no change in whether the painting was beautiful without there being a change in the instantiation of physical properties on the canvas.

A philosopher of art might agree or disagree with this supervenience thesis. Here are two ways in which they might disagree. First, they might disagree because they think context matters to whether or not a painting is beautiful. Whether a painting is beautiful isn't just determined by physical properties such as the pattern of paint on the canvas. It is also determined by the social context. If one thinks that the social context is also something physical, then one could reject this local supervenience of the aesthetic on the physical in favor of a global supervenience thesis: the aesthetic properties of things are not determined locally by their physical features alone, but globally by the physical features of the world as a whole. If this is true, then two paintings could be physically identical, while one is beautiful and the other is not, because they were created in different social contexts. But if two entire worlds are physically identical, then they must contain the same facts about which paintings in them are beautiful and which are not.

A second way one might reject the local supervenience of the aesthetic on the physical would be to reject supervenience physicalism altogether. One might think that the aesthetic properties of paintings are not determined by the patterns of paint on canvas, nor are they determined by the patterns of paint plus a broader social context. Rather, aesthetic features are irreducible, nonphysical features of our world.<sup>9</sup> It is an irreducible fact, not determined by anything physical, which paintings are beautiful and which are not. The aesthetic features of our world are something altogether independent of their physical features.

We entered this discussion by way of the mind-body problem. However, as we are seeing with this example, supervenience theses are debated much more widely in philosophy, not just in debates about the metaphysics of mind. Supervenience theses have been proposed as well in aesthetics, the philosophy of science (concerning the supervenience of laws, causal relations, and probabilities), the philosophy of language (concerning the supervenience of meanings), and moral philosophy (concerning the supervenience of moral and other normative features). In philosophy of mind, it is very common for metaphysicians to endorse the supervenience of

**Local supervenience:**  
the A-properties locally supervene on the B-properties just in case individuals that are indiscernible with respect to the instantiation of B-properties ("B-indiscernible," for short) are also indiscernible with respect to the instantiation of A-properties (A-indiscernible)

mental properties on physical properties, and to in this way explain the physical properties' relative fundamentality; what is usually debated is whether we should endorse something stronger than supervenience physicalism.

### EXERCISE 7.2

#### Supervenience

Many have debated whether the moral properties of an action supervene on its physical properties. Here are two supervenience theses one might consider:

(Local moral-physical supervenience) The physical properties of an action completely determine that action's moral properties. In other words, any two actions that are physical duplicates are moral duplicates.

(Global moral-physical supervenience) The physical properties of the world completely determine the moral properties of any action at that world. In other words, any two worlds that are physical duplicates are duplicates with respect to their moral properties.

Consider a situation in which one person slaps another across their face. Suppose this action is morally wrong. If so, what would be the most plausible supervenience basis for the wrongness of this action? Is your answer compatible with either supervenience thesis above? If so, which thesis and why?

## 7.4 The Limits of Supervenience

For many years, supervenience was the concept used to frame physicalist theses, a way to express the claim that although mental and other properties are not strictly identical with physical properties, they are, as it is frequently put, *nothing over and above* physical properties. They are, an ontological free lunch, and as such (it was frequently thought) their existence doesn't undermine physicalism. But is supervenience enough to guarantee physicalism? Today, a consensus view among metaphysicians is that the fact that some properties supervene on some others does not guarantee that the former are nothing over and above the latter. Actually, supervenience physicalism is compatible with dualism.

That this is so was noted already in the 1980s by Jaegwon Kim (1934-2019), the philosopher who did the most to rigorously formulate and distinguish a variety of supervenience theses and their usefulness

for stating metaphysical positions. In his 1984 paper, "Concepts of Supervenience," Kim noted that whether supervenience was a strong enough relation to guarantee physicalism depended on the variety of supervenience thesis adopted (see Text Box 7.3). Later, in the first edition of his *Philosophy of Mind* textbook, Kim noted that actually even the strongest versions of supervenience were not enough to ensure metaphysical dependence of the mental on the physical. There, Kim recommended that if physicalists wanted to state their view using a supervenience claim (that mental properties supervene on physical properties), then they should supplement the supervenience claim with a thesis of mind-body dependence. If one wants to be a physicalist, one should claim not only that mental properties supervene on physical properties, but also that mental properties ontologically depend on physical properties.

As Kim noted, a supervenience claim only ensures a kind of correlation between the instantiation of mental and physical properties. Even strong supervenience (see Text Box 7.3) of the kind endorsed by most physicalists entails only a necessary correlation between mental and physical properties. But there are many ways that such a correlation between mental and physical properties might obtain, and some of these ways are incompatible with physicalism. For example:

- Mental properties could be distinct from physical properties. There is a necessary correlation between mental and physical properties, but this is just a brute fact.
- Mental properties could be distinct from physical properties. However, there is an omnipotent (all-powerful) God that ensures that whenever some particular physical properties are instantiated, some particular mental properties are also instantiated.
- Mental properties could be distinct from physical properties. There is a necessary correlation between mental and physical properties because there are psychophysical laws connecting the instantiation of physical properties with the instantiation of mental properties.

This second metaphysical hypothesis was proposed by the early modern philosopher Gottfried Leibniz (1646-1716). Leibniz, like Descartes, thought that minds and bodies were distinct. But he couldn't see how it could be that immaterial minds could causally interact with material bodies. His solution was to propose a thesis of **preestablished harmony**. There is a harmony between the properties we observe in bodies and those we introspect in our minds not because the two causally interact as Descartes thought, but because God sustains a systematic correlation between the two. If God is omnipotent, He can make it the case that there is no change in the mental properties without a corresponding change in the physical properties. But this is not because some form of physicalism is true. It is

**Preestablished harmony:** mental and physical properties are correlated not because they ontologically depend on one another, but because the correlation is set up that way by God

because God makes the supervenience claim true. Indeed, it seems, God could even make it that a strong supervenience claim was true and that the physical properties were *necessarily* correlated with the mental properties. The third hypothesis is one that was discussed and defended by the philosopher David Chalmers in his 1996 book *The Conscious Mind*. Since physical and mental properties are distinct on this view, it is a version of dualism. Chalmers considers it a “naturalistic” dualism since the instantiation of mental properties arises out of the instantiation of physical properties and the laws of nature. In his book, Chalmers does not argue that the laws of nature are necessary, but other philosophers have held this position (e.g., Swoyer 1982, Shoemaker 1998). If so, we arrive a version of naturalistic dualism compatible with strong supervenience. The coherence of each of these hypotheses shows that even strong supervenience is compatible with dualism, and thus supervenience theses in general are inadequate to formulate physicalism.

The metaphysician Jonathan Schaffer (2009) notes that we should have realized already that supervenience would fail as a way of formulating metaphysical positions like physicalism, because it is a relation that lacks the correct formal features. For example, to capture the relation between the fundamental physical properties and all others, we really want a relation that is **irreflexive**. That is, we want a relation that nothing can bear to itself. This is because it is part of the meaning of ‘fundamental’ that nothing can be more fundamental than itself. But, supervenience is a **reflexive** relation. For any set of A-properties, the A-properties supervene on the A-properties.

Second, if we want to characterize the relation between a set of properties that are more fundamental (like the physical properties) and a set that are less fundamental (like the mental properties), then we really need a relation that is **asymmetric**. This is because if the A-properties are more fundamental than the B-properties, then it can’t also be the case that the B-properties are more fundamental than the A-properties. But the supervenience relation fails to be asymmetric. We see this in the case of preestablished harmony. If our world was that way, then it would be the case both that the mental properties supervened on the physical properties and the physical properties supervened on the mental properties. So, Schaffer argues, the relation the physicalist wants is one that is irreflexive and asymmetric. This is a view that is widely shared.<sup>10</sup>

Here is where the concept of grounding is supposed to come to the rescue. Grounding, unlike supervenience, Schaffer claims, provides a characterization of the relationship between the fundamental physical properties and all others that is incompatible with dualism and that has the correct formal properties. Grounding is irreflexive and asymmetric. (It is also **transitive**.) It is able to describe how a certain class of properties

**Irreflexivity:** a relation R is irreflexive iff for any x it is false that x bears R to itself

**Reflexivity:** a relation R is reflexive iff for all x, x bears R to itself

**Asymmetry:** a relation R is asymmetric iff for all x and y, if x bears R to y, then y does not bear R to x

**Transitivity:** a relation R is transitive iff for all x, y, and z, if x bears R to y, and y bears R to z, then x bears R to z

may be more fundamental than all others, and is stronger than a mere correlation (even necessary correlation) between two sets of properties.

### EXERCISE 7.3

#### Socrates and singleton Socrates

In set theory, given any object that exists, there necessarily also exists the set that contains as a member only that object. For example, if Socrates exists, then it will also be the case that the set whose only member is Socrates exists. This singleton set, which we may call 'singleton Socrates' is represented as: {Socrates}.

Say that the B-facts supervene on the A-facts just in case there can be no change in the B-facts without change in the A-facts.

1. Is it the case, given set theory, that facts about the existence of singleton Socrates supervene on facts about the instantiation of Socrates?
2. Is it the case, given set theory, that facts about the existence of Socrates supervene on facts about the existence of singleton Socrates?
3. Which entity is more fundamental: Socrates or singleton Socrates?
4. Do your answers to the previous questions conflict with the following principle:

For all  $x$  and all  $y$ ,  $x$  is more fundamental than  $y$  iff the facts about  $y$  supervene on the facts about  $x$ .

## 7.5 Grounding

Many metaphysicians working today agree that what is at stake in metaphysical debates like the mind-body problem is not whether the instantiation of some set of properties correlates with the instantiation of properties in some other set, but something else. This something else may be expressed in different ways:

- whether the instantiation of one set of properties *depends on* the instantiation of the other set of properties, or
- whether one set of properties is instantiated *in virtue of* the instantiation of the other set of properties, or
- whether the instantiation of one set of properties is *nothing over and above* the instantiation of the other set of properties.<sup>11</sup>



**Ground:** a relation between facts or entities that makes for ontological dependence

**Grounding** is a conceptual tool that was introduced in order to capture these ideas of what is really at stake in such metaphysical debates.

In an influential paper from 2009, “On What Grounds What,” Schaffer argued that many of the most significant issues debated by metaphysicians are really questions about grounding. These include the mind-body problem, but also the questions of realism or nominalism about properties and other abstract entities, the existence of God, and the debates about time we will discuss in the next chapter.<sup>12</sup> The question in all of these cases is not what exists, but rather what grounds what. Do physical properties ground mental properties? Are properties grounded in sets of their instances? Is there a God who is a supernatural being or something that is merely grounded in our beliefs and scriptures? Are facts about the past and future grounded in events that occur earlier or later than the present moment or just people’s memories and expectations?

In the decade that followed, metaphysicians grappled with Schaffer’s proposal. On the one hand, most agreed that many of the most significant metaphysical debates are not just about what exists. Often, as in the case of the mind-body problem, all parties agree about what exists. The question is what is more fundamental, and whether one kind of entity depends on or is there in virtue of or is nothing over and above the other. But two kinds of resistance arose.

The first kind of resistance to Schaffer’s proposal came from philosophers who insisted they could not make clear sense of this notion of grounding. Papers by Chris Daly (2012) and Thomas Hofweber (2009), for example, argued that the notion of grounding was obscure. Although Schaffer characterized grounding as a relation that was asymmetric, irreflexive, transitive, and sufficed for facts about relative fundamentality to obtain, he also insisted one could not provide grounding with a definition in terms of other, antecedently understood concepts. Unlike supervenience, which was introduced into debates about the mind-body problem by Kim and others by way of precise and intelligible definitions (see Text Box 7.3), Schaffer introduced grounding as a **primitive concept**. Grounding skeptics suggested we instead stick to using notions for which we can provide clear and precise definitions.

**Primitive concept:** a concept is primitive if it does not have an analysis in terms of other concepts

Other philosophers didn’t object to the terminology of grounding, but rather questioned Schaffer’s claim that we should use the term to denote a unique and special metaphysical relation. Instead, for example, Jessica Wilson (2014) argued that yes, it is grounding, not supervenience that is really the issue in many central metaphysical debates. But the word ‘grounding’ doesn’t denote one single relation (what she refers to as “Big-G Grounding”) but rather a rich variety of “little-g” grounding relations. The little-g grounding relations include all of those metaphysical relations that make for nothing-over-and-aboveness, including identity, mereological composition, realization (see Text Box 7.2), and set-formation. Kathrin

Koslicki (2015) pointed out that lumping these relations all under a single concept was to take a step backward, rather than forward, in coming to understand the complex and diverse metaphysical relations that structure our world. Karen Bennett (2017) agreed, though she calls the relevant class of relations “building relations” rather than grounding relations. What ties together all of the various building relations, according to Bennett, is that they are all asymmetric, imply a necessitation relation between facts or entities, and involve one set of entities in some sense generating another.<sup>13</sup> But, Bennett notes, just because all building relations share these characteristics, this doesn’t imply there is just one building relation. Mereological composition, set-formation, and realization, for example, are very different ways of bringing things together to create something new. Moreover, we learn more about the structure of the world by carefully distinguishing different such relations.

Schaffer has addressed these challenges in his work. Indeed, to those who find the concept obscure, he has always argued that grounding is a notion that is familiar and has been well-understood in philosophy since at least Plato. He points to Plato’s dialogue “Euthyphro” whose central question is a question about the direction of a grounding relation: Is what is pious so because it is what the gods love? Or do the gods love what is pious because it is pious? The question here is whether the facts about piety ground the facts about what is loved by the gods, or whether the facts about what is loved by the gods ground the facts about what is pious. Schaffer grants that metaphysicians raised in the Quinean tradition (see Chapter 1) may be more familiar and comfortable framing all metaphysical questions as ontological questions, but he presses:

Of course the notion of grounding may be *unfamiliar* to some metaphysicians raised only on Quine and Carnap. The best advice I can give is *work with the notion*, and see if you then come to grasp it. (2009, p. 376)

It is fair to say that many metaphysicians have taken his invitation. The use of the notion of grounding in metaphysics has exploded over the last decade.

The debate about the proper way to understand grounding is still young and so it is too soon to tell whether Schaffer’s “big-G” grounding concept is here to stay. What we can say at this time is that research into grounding in its many forms has yielded a wealth of tools for metaphysicians engaged in first-order projects.<sup>14</sup> Philosophers working on, for example, the mind-body problem or nominalism or the philosophy of time, can pick and choose which notions are most useful for their purposes, which best capture the metaphysical relationships in their domain of interest. They can then decide whether this is the “big-G” grounding favored by Schaffer, or for example, one of the many “little-g” grounding or building relations

avored by Wilson and Bennett. Some, like Schaffer will want to characterize physicalism as the claim that everything is physical or is grounded in the physical, in the “big-G” sense. Some, like Andrew Melnyk and Jessica Wilson, will want to characterize physicalism as the claim that all properties are physical properties or realized by physical properties. Others, like Thomas Polger and Lawrence Shapiro (2016), argue that we should go back to understanding the relation between the mental and physical as one of identity. The correct relation to use to characterize the relative fundamentality of the physical vis-à-vis the mental is a topic of ongoing discussion in the metaphysics of mind.

#### EXERCISE 7.4

##### Identity and Grounding

Here is an argument that identity is not a grounding relation. Which if any premises could someone like Wilson question in order to claim the argument is unsound?

1. Suppose  $a = b$ .
2. Suppose identity is a grounding relation.
3. Then (it follows from (1) and (2)),  $b$  grounds  $a$ .
4. But, identity is a symmetric relation:  $\forall x \forall y (x = y \supset y = x)$ .
5. Therefore (from (1) and (4)),  $b = a$ .
6. Then, (from (2) and (5)),  $a$  grounds  $b$ .
7.  $\forall x \forall y (x \text{ grounds } y \supset x \text{ is more fundamental than } y)$
8. Then (from (3) and (7)),  $b$  is more fundamental than  $a$ .
9. And (from (6) and (7)),  $a$  is more fundamental than  $b$ .
10. But the *more fundamental than* relation is asymmetric:  $\forall x \forall y (x \text{ is more fundamental than } y \supset \neg (y \text{ is more fundamental than } x))$ .
11. Therefore (from (8) and (10)),  $\neg (a \text{ is more fundamental than } b)$ . CONTRADICTION with (9).
12. Therefore, by reductio, identity is not a grounding relation.

## 7.6 Fundamentality

As we’ve seen, many central metaphysical controversies are not primarily debates about existence, but rather about grounding or relative fundamentality. For example, contemporary discussions of the mind-body problem concern the question of whether mental properties ultimately depend on physical properties that are more fundamental, and if so, then how do they depend on these more fundamental physical properties.

We are now at a place where we may ask the question of what one could mean by saying that physical properties are fundamental, or at least, are more fundamental than mental properties. There are many ways in which metaphysicians have understood fundamentality talk, indeed many more ways than we will have the chance to discuss here. In this and the following two sections, we will discuss the primary, most influential ways of understanding what it means to say that something is fundamental.

Before we get started, it is worth distinguishing the sense of 'fundamental' with which we will be interested from other uses of the term. For example, in many contexts, one might use 'fundamental' to mean what is most important or central. But when physicalists say physical entities are fundamental, they are not saying that physical entities are more central or more important than the entities of (say) other sciences. They are in the first place making a metaphysical claim, not a normative one. What this metaphysical claim comes to is the topic of the next two sections.

We can also distinguish metaphysical fundamentality from notions of *epistemological* fundamentality. When we discuss epistemological fundamentality, we are after what it is that is in the first place justified or what primarily we may have knowledge of. For example, in epistemology, there is a historically influential position called 'foundationalism.' This is the view that there are a class of epistemologically basic items that provide a justificatory foundation for all of the rest of our beliefs about the world. These are usually thought to be beliefs about our sensory experiences. Foundationalists argue that beliefs about our sensory experiences don't require any further justification because we cannot be wrong about what experiences we are having. This is an interesting claim, but we should notice that it is different from the claim that our beliefs about our sensory experiences form a *metaphysical* foundation for everything else. That would be to say that everything (or every other belief) is ultimately made up out of sensory beliefs (a view that is very different and sounds like idealism). So, claiming something is epistemologically fundamental is very different from claiming it is metaphysically fundamental. We must keep these two notions of fundamentality separate.

Note finally that metaphysicians like to distinguish the notions of absolute and relative fundamentality. The absolutely fundamental entities, if any exist, would be those that are fundamental full stop. As we will discuss in Section 7.9, it is an open question whether or not there are any entities that are absolutely fundamental. This makes it useful to have notions of relative fundamentality on hand. When we talk about relative fundamentality, we are talking about which entities are more or less fundamental than which others and which entities are equally fundamental as which others. Most metaphysicians believe that even if there is no absolutely fundamental level to reality, there can be still be facts about relative fundamentality. For example, even if there will be no end to physics's search

for the ultimate building blocks of nature, it can still be a fact that electrons are more fundamental than elephants.

## 7.7 Fundamentality as Ontological Independence

The first way in which metaphysicians will often explicate the notion of fundamentality is in terms of ontological independence. That is to say, the absolutely fundamental entities are those that are not dependent on any other entities, in an ontological sense. They do not depend on any other entities for their existence or natures.

Ontological independence is a common way metaphysicians understand fundamentality talk today. It is the way the notions are explicated in influential works, for example, by Karen Bennett (2017) and Jonathan Schaffer (2009). However, this understanding does have some limitations and complications. First, what is ontological independence? There are a variety of notions one could have in mind here. The grounding framework provides one. Indeed, Schaffer's (2009) proposal is to understand the absolutely fundamental entities as those that are ontologically independent in the sense of being ungrounded. But if one is skeptical that there is such a thing as grounding, then what could it mean to say that the absolutely fundamental entities are those that are ontologically independent?

A natural place to start is to understand ontological independence as a kind of **modal independence** or as Jennifer Wang (2016) puts it, modal freedom. Chapter 10 discusses the topic of modality in detail. Modality concerns what is possible or impossible, what is contingent or necessary. So, as a first pass, we may say that an entity is modally independent just in case its existence does not necessitate the existence of any other entity.

**Modal independence:**  
an entity is modally independent just in case its existence does not necessitate the existence of another entity

Modal independence (weak version):  $x$  is modally independent iff  $\forall y$   
 $(y \neq x \supset \Diamond (\exists z (z = x) \wedge \neg \exists w (w = y)))$ ,

where the ' $\Diamond$ ' symbol stands for "it is possible that." In English, one can read this as:  $x$  is modally independent if and only if for any  $y$  (distinct from  $x$ ), it is possible that  $x$  exists but  $y$  does not. Or one could make the even stronger claim that a modally independent entity would be an entity that could possibly exist all on its own, without any other:

Modal independence (strong version):  $x$  is modally independent iff  $\Diamond$   
 $(\exists z (z = x) \wedge \forall y (y \neq x \supset \neg \exists w (w = y)))$

This is equivalent to saying that the modally independent entity's existence doesn't necessitate that of any other entity at all.

In some ways, the analysis of ontological independence as some kind of modal independence seems right. For example, many physicalists think mental properties fail to be fundamental because their instantiation

depends on the instantiation of other properties that are physical (e.g., neurophysiological properties). If one understands ontological independence as modal independence, then this is to say that physicalists think mental properties fail to be ontologically independent because their existence necessitates the existence of some physical properties. There is no possibility of having mental properties instantiated without at least some physical properties instantiated.

Physicalists also typically believe that there are some physical properties that are absolutely fundamental. If the physicalist then understands absolute fundamentality as ontological independence, and ontological independence as the second, stronger kind of modal independence stated above, then this is to say it is possible for there to be a world in which only physical properties are instantiated, where each is independent of all of the others. The physical properties (or at least the most fundamental physical properties) in a way “stand on their own” and don’t require any other properties for their instantiation.<sup>15</sup>

There are a couple of immediate problems with understanding the kind of ontological independence we are after in terms of modal independence. And this is why most metaphysicians, Wang included, think it is obvious that ontological independence is something distinct from modal independence. First, physicalists will face a tension with two claims they want to make: (i) that there is a class of physical entities that are ontologically independent, and (ii) that everything else, for example, mental entities, supervene on these physical entities. And this is because, as we have seen, the kind of supervenience claim physicalists generally want to make is a strong supervenience claim, one that says that the instantiation of physical properties necessitates (or determines) the instantiation of all other properties at our world. But if the instantiation of the physical properties necessitates the instantiation of e.g., mental properties, then this is to say you cannot have the physical properties without the mental properties. And this is to say that the physical properties are not modally independent in the sense defined above.<sup>16</sup>

Now one can fix this by modifying the modal independence requirement to say that a modally independent entity’s existence doesn’t necessitate the existence of any other *fundamental* entity. But this will not do if we are trying to use the concept of modal independence to analyze or define the concept of fundamentality. It would make that definition circular. Anyway, as Wang argues, this won’t help too much since even the fundamental entities seem to necessitate the existence of each other. For example, many if not most metaphysicians hold an ontology that contains both particular objects and properties. And one can’t have an object without properties, nor can there be free-floating properties that are not instantiated by any objects.<sup>17</sup> So, fundamental entities of different ontological categories appear to necessitate each other’s existence.

Another issue arises for those who believe there are entities that necessarily exist. For example, in Chapter 2, we discussed mathematical Platonism: the view that there exist numbers, sets, and other mathematical objects. Platonists often hold the view that numbers, and at least the pure sets (sets whose members do not involve contingently existing objects) are necessary existents. Many theists, those who believe in God, take Him to be a necessary existent as well. But if either of these types of views are correct and there are some necessary existents, then this would mean nothing (or nothing else) is modally independent in the senses defined above. There are no objects that can exist without these necessary existents.

What all of this suggests is that modal independence is not a good way to capture the kind of ontological independence metaphysicians are after when they are trying to understand fundamentality. It isn't even true let alone analytic that the fundamental entities are those that are modally independent. So, we should try to find some other way to regiment ontological independence talk.

One option we have already mentioned is to understand ontological independence in terms of ("big-G") grounding, as Schaffer does. The fundamental entities are those that are ungrounded. If one is skeptical that there is one single grounding relation, then one could also understand ontological independence by appeal to what Wilson calls the ("little-g") grounding relations and what Bennett calls the building relations: mereological (part/whole) relations, realization, set formation, etc. For example, one might argue that ontological independence should be understood as mereological independence:

$x$  is ontologically independent iff  $\neg\exists y (y \neq x \wedge y \text{ is a part of } x)$ .

In other words, one might claim that the ontologically independent entities are those that lack any proper parts - they are the mereological atoms or simples discussed in Chapter 3.

### EXERCISE 7.5

#### Fundamentality and Ontological Independence

For each of the following entities, in your view are they (i) absolutely fundamental, (ii) ontologically independent in the sense of being ungrounded, (iii) modally independent in the weak sense, and/or (iv) modally independent in the strong sense? Explain each answer.

- a. Electrons
- b. God
- c. The number 9

**TEXT BOX 7.4****Fundamentality as a Mereological Notion**

In a classic paper in the philosophy of science from 1958, “Unity of Science as a Working Hypothesis,” Paul Oppenheim (1885-1977) and Hilary Putnam (1926-2016) conjectured that there are levels to reality that correspond to the divisions in the sciences. This includes a most basic and fundamental level to reality that consists of the entities postulated by physics. These entities are fundamental in the sense that they are mereologically simple. The entities of all other sciences are all less fundamental in the sense that they are completely decomposable into entities in the domain of physics. The entities at the level of chemistry are directly composed out of the entities of physics. The entities at the level of biology are directly composed out of the entities of chemistry, and so indirectly composed out of the entities of physics. And so on, up the chain of being.

Although it is natural to want to understand fundamentality as a kind of mereological independence, there has been resistance to Oppenheim and Putnam’s proposal in the metaphysics and (even more so) philosophy of science literature. There are a couple of recurring concerns. First, philosophers have questioned whether the sciences can be neatly carved up into levels in the way that Oppenheim and Putnam’s model assumes. Probably the most persuasive arguments have come from philosophers of neuroscience (e.g., Craver 2007). In neuroscience, explanations typically span across Oppenheim and Putnam’s ontological levels. A typical explanation in neuroscience will appeal to brain regions, cells (neurons), molecules, and ions. This problematizes the conception of the sciences as stratified in this way. Second, even if we

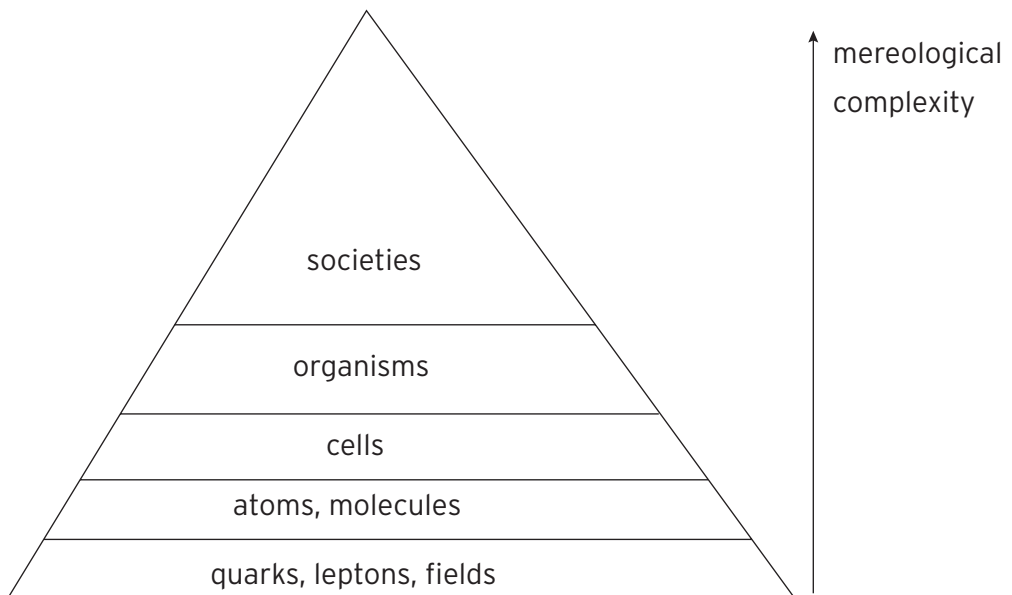


Figure 7.1 Levels of Reality



allow the assumption that there are levels of reality corresponding to the different sciences with physics at the bottom, philosophers have questioned whether it is accurate to think of the metaphysical relationships between entities at different levels as uniformly that of mereological composition. We are very familiar with pictorial representations of atoms that depict them as mereological composites of subatomic particles neatly separated in space, or of cells as neatly built up out of chemical components – the nucleus in the middle, the mitochondria off to the side. But do these pictorial representations accurately depict metaphysical structure?<sup>18</sup> Moreover, as we have seen, philosophers of mind do not typically use mereological notions to understand the relationship between mind and body (or mental and physical properties), but rather make use of notions like supervenience and realization. So, at the very least, Oppenheim and Putnam’s model would have to be broadened to take into account a richer variety of “little-g” grounding relations. Despite these two concerns, the picture remains influential and even if it does not do an accurate job of representing the whole of reality, one might think that it gets right at least some of the relationships between more and less fundamental entities that actually exist, where mereological relationships between entities in the domain of different sciences obtain.

There are two issues that arise for anyone who holds the view that fundamentality is ontological independence (whether ontological independence is understood in terms of grounding or in some other way). The first concerns whether this picture is too restrictive about the way the world could fundamentally turn out. Isn’t it conceivable that the world is fundamentally such that there exist a collection of entities that are mutually dependent on one another? For example, it seems possible that the fundamental particles ontologically depend on the fundamental fields, and the fundamental fields ontologically depend on the fundamental particles. In a world like this, neither particles nor fields are independent, but both are still absolutely fundamental. Or to cite another example, some philosophers of science believe that there are mutual dependence relations holding between the fundamental laws of nature and the fundamental properties. If either of these views is conceivable, then the view that fundamentality just is ontological independence would seem to fail, because there could be fundamental entities that ontologically depend on each other.

A second issue concerns how this view looks when paired with the Bennett/Wilson position that there are multiple grounding or building relations. For example, consider the “little-g” grounding relations

mereological composition and realization. If we believe that all grounding relations make for relative fundamentality, then we have two principles:

If  $x$  is a part of  $y$ , then  $x$  is more fundamental than  $y$ .

If  $x$  realizes  $y$ , then  $x$  is more fundamental than  $y$ .

So far so good. But what happens if it turns out for some  $x$  and some  $y$  that both  $x$  is a part of  $y$  and  $y$  realizes  $x$ ? Then it appears that both:

$x$  is more fundamental than  $y$

and

$y$  is more fundamental than  $x$ .<sup>19</sup>

This looks problematic because the ‘is more fundamental than’ relation is asymmetric. This means it can’t be the case for some  $x$  and some  $y$  that  $x$  is more fundamental than  $y$  and  $y$  is more fundamental than  $x$ .

An option that Bennett considers and endorses in her 2017 book is to allow that there are multiple senses of fundamentality corresponding to the different building relations. So, in the case we are imagining, there are really two distinct fundamentality relations (and notions of absolute fundamentality) in play: fundamentality<sub>composition</sub> and fundamentality<sub>realization</sub>. And then what is true is:

$x$  is more fundamental<sub>composition</sub> than  $y$

and

$y$  is more fundamental<sub>realization</sub> than  $x$ .

There is no tension between these claims and the fact that each individual relative fundamentality relation is asymmetric.

This is interesting, but does raise the question: if there are multiple properties of absolute fundamentality (fundamentality<sub>composition</sub>, fundamentality<sub>realization</sub>, fundamentality<sub>set formation</sub>, etc.), does this mean that there is no single absolute fact about what are the basic entities at our world? When trying to explicate the notion of absolute fundamentality, metaphysicians often like to use the “all God would have to do to create the world” metaphor. The fundamental entities are such that all God would have to do when creating the world is create those entities. Everything else would come for free. But if there are multiple fundamentality properties, then what is it that is all God would have to do to create the world? Would he just have to create the things that are fundamental in one of these senses? Or all of them? But surely, that can’t be right if the things that are fundamental in one sense automatically bring about those that are fundamental in another of the senses. One gets the feeling that really there is a privileged notion of fundamentality among the many, which brings us

back to Schaffer's view and his proposal that there is one notion of fundamentality, that of being ungrounded (in the "big-G" sense).

### TEXT BOX 7.5

#### Grounding and Fundamentality

In her work, Jessica Wilson calls relations like the part/whole relation, realization, and even identity "little-g" grounding relations. This is because if  $x$  (or the  $x$ s) ground  $y$  in any of these senses, then  $x$  (or the  $x$ s) are more fundamental than  $y$ . However, it is controversial whether these are all truly grounding relations, in the sense of making for relative fundamentality of the grounding base vis-à-vis the grounded.

Exercise 7.4 discussed an argument for the conclusion that identity is not a grounding relation. There is reason as well to question whether the parthood relation is a grounding relation. It is natural to think that the following principle is true:

If  $x$  is a part of  $y$ , then  $x$  grounds  $y$ ,

because it also seems true that:

If  $x$  is a part of  $y$ , then  $x$  is more fundamental than  $y$ .

But influential papers by Jonathan Schaffer<sup>20</sup> have argued that these principles are not universally valid. For both assume a metaphysical position he calls **priority pluralism**.

Priority pluralism: parts are more fundamental than wholes.

However, one could also hold a different position:

**Priority monism:** the one ultimate whole is more fundamental than its parts.<sup>21</sup>

In his work, Schaffer develops several arguments in favor of priority monism. But for the issue we are considering here, it doesn't really matter whether priority pluralism or priority monism is correct. What matters is simply that both positions are coherent. If it is even conceivable that our world could be such that wholes are more fundamental than their parts, then this means that the parthood relation does not automatically make for relative fundamentality, and so parthood is not a grounding relation. Protons, neutrons, and electrons could mereologically compose an atom, but this would

**Priority pluralism:** the view that parts are more fundamental than wholes

**Priority monism:** the view that the one ultimate whole is more fundamental than its parts

not entail that the atom is grounded in those protons, neutrons, and electrons. It could instead be the case that relative fundamentality goes in the opposite direction and that the protons, neutrons, and electrons are grounded in the atom.

Why should someone think that wholes are more fundamental than their parts? One of Schaffer's arguments is relatively technical and comes from a consideration of quantum mechanics.<sup>22</sup> However, we can discuss an *a priori* argument that relates back to our discussions of mereology in Chapter 3. Remember there we considered the possibility of gunk. The world is gunky just in case there are ultimately no mereological atoms, and every object, even those of our most fundamental physical theories, have proper parts. Here is one statement of Schaffer's argument:

1. For all we know, the world is gunky.
2. Then (by (1)), for all we know, there are no mereological atoms.
3. If priority pluralism is true, then the fundamental entities are the mereological atoms.
4. So (by (2) and (3)), if priority pluralism is true, then for all we know, there are no absolutely fundamental entities.
5. We know there must be some absolutely fundamental entities.
6. Therefore (by (4) and (5)), priority pluralism is not true.
7. Therefore (from (6)), the wholes are more fundamental than the parts.

The epistemological possibility of gunk shows that priority pluralism must be false.

At this stage, one might ask, couldn't one run an argument in the opposite direction against priority monism? For, consider the possibility of junk. The world is junky just in case every entity is a proper part of something. If the world is junky, then there might be mereological atoms, but there wouldn't be what we might call "ultimate wholes": wholes that aren't in turn proper parts of anything else. If junk were epistemologically possible, then we could try to run the following argument:

1. For all we know, the world is junky.
2. Then (by (1)), for all we know there are no ultimate wholes.
3. If priority monism is true, then the fundamental entity is an ultimate whole.

4. So (by (2) and (3)), if priority monism is true, then for all we know, there are no absolutely fundamental entities.
5. We know there must be some absolutely fundamental entities.
6. Therefore (by (4) and (5)), priority monism is not true.
7. Therefore (from (6)), parts are more fundamental than wholes.

The trouble with this argument, as Schaffer points out, is that unlike gunk, junk is not epistemologically possible. Classical mereology implies that there is a unique sum for each set of objects (Schaffer 2010, p. 64). And since in the cases of mereological composition we are considering, a whole spatially overlaps its parts, there will not be a mereological sum of the whole with its parts. So whatever objects exist in our world, there will be what we are calling an “ultimate whole.” Priority monism is safe from the argument from junk.

## 7.8 The Fundamental as a Basis for the Nonfundamental

So, one influential way of understanding what it means to say something is fundamental is to say that it is ontologically independent or ungrounded. However, there are some issues that arise for the view, including the concern that some of the absolutely fundamental entities might be ontologically dependent on each other. As an alternative, one could argue that the concept of fundamentality (either absolute or relative fundamentality) is primitive. There is just a brute set of facts about which entities are absolutely fundamental and which are more fundamental than which others, and there is no more basic analysis one can give of fundamentality talk. This is a view that certainly has some adherents (Fine 2001, Sider 2011, Wilson 2014). However, another strategy is to try to define absolute fundamentality in terms of something other than ontological independence. In this section, we will consider defining the absolutely fundamental in terms of the ability of a class of entities as a whole to provide a metaphysical basis for everything else.

According to a standard implementation of this strategy, the absolutely fundamental entities are those that together provide a metaphysical basis for the nonfundamental that is:

Complete: All entities are either fundamental entities or based in the fundamental entities.

And, depending on how one understands ‘basing,’ the account may need to add an additional principle, namely, that the metaphysical basis for the nonfundamental is:

Minimal: It is not the case that all entities are based in any proper subset of the fundamental entities.

The second clause would be needed in order not to trivialize the concept of the fundamental, if one understood ‘basing’ in such a way that entities could be based in themselves. If the only requirement on being fundamental was being a member of a set of entities in which all are based, and it is possible for entities to be based in themselves, then any entity could count as fundamental since it is a member of the set of all entities. (If one understands ‘basing’ as big-G grounding, then one doesn’t need the Minimal thesis to characterize the fundamental, since big-G grounding is an irreflexive relation. However, minimality is required if one understanding ‘basing’ in terms of supervenience, which is a reflexive relation.)

One might ask what it means to say that all entities are “based in” the fundamental entities. Here, metaphysicians will vary depending on their preferred view about grounding and other metaphysical relations. Some will interpret “All entities are based in the fundamental entities,” to mean all entities are grounded in the fundamental entities. Others will understand ‘basing’ in terms of supervenience or realization.

Alternatively, some metaphysicians prefer to adopt an interpretation according to which the relevant sense of ‘basing’ is explanatory. In this sense, a set of entities is not fundamental unless one could use that base to provide a complete explanation of the existence and nature of all other entities. This brings the metaphysical notion of fundamentality closer to the sense one traditionally finds in the philosophy of science literature on reduction. Reduction has historically been understood as an explanatory notion.<sup>23</sup> One reduces *x* to *y* just in case one can provide a complete explanation of *x* just in terms of *y*. The idea that the fundamental is a reductive or explanatory basis for all else is still prevalent in metaphysics and philosophy of science today (e.g., Sider 2011, Chalmers 2012, Ney 2021b). But it is also very controversial (e.g., Block and Stalnaker 1999, Schaffer 2017). For some would argue that the fundamental is purely a metaphysical notion: it has to do simply with the nature of reality and nothing to do with our epistemological relationship to it. Even if we cannot use the fundamental to explain everything else, perhaps because of our conceptual or computational limitations, the fundamental still plays its role in being the metaphysical basis or ground for all else. Thus, whether explanatory notions should be used to characterize fundamentality is contested.

Coming back to the more general idea of the absolutely fundamental as a complete (perhaps, minimal) basis, one thing to like about this proposal is that it seems to avoid one of the worries noted at the end

of the last section for construals of fundamentality as ontological independence. Understanding the fundamental as a basis for everything else is consistent with the possibility that at least some of the fundamental entities depend on each other (for example, the particles and the fields or the properties and the laws). So long as the basis for everything else in the world includes both of them, entities could be interdependent and fundamental.

On the other hand, we also raised the question of what happens if there are multiple grounding or building relations. Would this imply that there is no single, objective fact about which entities are absolutely fundamental? A similar issue arises for the basis interpretation of absolute fundamentality. What if it turns out that two sets of entities X and Y form two distinct complete bases? We noted at the end of the previous section that this does seem to create a tension with the “all God had to do” metaphor that metaphysicians use to motivate some discussions of fundamentality. But is such a metaphor essential to our way of understanding fundamentality? It is not clear. In philosophy of science, there have certainly been “ontological pluralists” who are happy to allow multiple, non-competing sets of fundamental entities (most famously, Dupre 1993). The renewed interest in fundamentality on the part of metaphysicians is still young and so there is much work to be done in hashing these questions out.

### **EXERCISE 7.6**

#### **Multiple Bases?**

In your view, is it conceivable that there be multiple, distinct sets of entities that are both absolutely fundamental? Why or why not? If your answer is yes, then what would be an example?

### **TEXT BOX 7.6**

#### **The Truthmakers as a Complete Basis**

In this chapter, we have focused on relative fundamentality relations between entities (such as Schaffer’s notion of grounding). However, another influential approach to fundamentality relies on the truthmaking framework introduced in Chapter 2. Here, the truthmakers are understood to be a fundamental class of entities. But truthmaking is not necessarily a relation between entities. It is a relation between entities (truthmakers) and truths. Truthmakers are fundamental in a sense related to what was discussed in Section 7.8: they are a complete basis

for all of the truths. This alternative approach to thinking about fundamentality has been developed in work by David Armstrong (2004) and John Heil (2012), among others.<sup>24</sup>

One difference between the grounding framework of Schaffer's discussed above and the truthmaking approach is that the former involves a layered conception of reality, while the latter's conception of reality is flat. For Schaffer, there are layers of entities bearing grounding relations to each other. These ultimately all bottom out in a fundamental level of what is ungrounded. By contrast, truthmaker theorists like Heil argue that there is just one level to reality consisting of all of the truthmakers. This one level is capable on its own of providing truthmakers for all of the truths. It is controversial among truthmaker theorists what this ultimate reality consists in: objects bearing properties and relations (Armstrong) or objects bearing intrinsic properties only (Heil). But this picture of reality as just consisting of the ultimate truthmakers can be motivated by its ontological parsimony. So long as there are truthmakers for all of the truths, truthmaker theorists don't think we have to be ontologically committed to any nonfundamental entities. For example, consider truths like "Some apples are red" or "Many baseball pitchers are left-handed." Grounding theorists typically believe in entities like apples and baseball pitchers. They don't think these things are fundamental, but rather they are nonfundamental entities ultimately grounded in whatever is fundamental. The truthmaker theorist though typically won't have an ontology including apples or baseball pitchers. They will have some fundamental ontology of truthmakers that are capable of making true claims like "Some apples are red," or "Many baseball pitchers are left-handed," but their ontology won't necessarily include apples or baseball pitchers themselves. And so, there is the thought that the truthmaker theorist's ontology is more parsimonious. Or so the argument goes.<sup>25</sup>

## 7.9 Is There a Fundamental Level?

Many projects in metaphysics (and science) assume that there is a fundamental level to reality, and that it is part of our task as philosophers and scientists to discover it. Whether absolute fundamentality is understood in terms of ontological independence, a basis for everything else, or taken as a primitive, there is this idea that there are some ultimate thing(s) that exist and that form the basis for everything else we interact with and see. But is there justification for thinking that there is a fundamental level to reality, a set of entities that are absolutely fundamental? Or is reality infinitely complex in such a way that there is no metaphysical foundation and there is only a forever descending chain of entities that are more and more fundamental?

It has proven very difficult to find a compelling argument for the claim that reality bottoms out in a class of absolutely fundamental entities (a fundamental level of reality). As Tuomas Tahko (2018) notes, this is more often taken as an axiom by philosophers and scientists – a starting



point to guide inquiry – not something that is argued for on the basis of other premises.

In his 2011 book, *Writing the Book of the World*, Theodore Sider suggested that the assumption of a fundamental level isn't really problematic in that there doesn't seem to be any good argument against it. After all, what would the argument against a fundamental level to reality look like? It probably wouldn't be a **deductive argument**, since the idea of a fundamental level to reality is at least conceivable. So it seems like it would have to be an **inductive argument**, something like:

Physicists once thought that everything depended on the features of molecules. But molecules gave way to atoms, which gave way to protons, neutrons and electrons, which have given way to quarks, leptons, and gauge bosons. Each time a new particle was discovered, physicists posited new features of the newly discovered particles, whose distribution accounted for, but could not be accounted for in terms of, the distribution of the distinctive features of the older particles. This historical progression of theories will probably continue forever, so there are no ultimate features on which everything depends. (2011, p. 135)

But, Sider argues, this is not a very compelling inductive argument, for several reasons.

Consider first what many take to be the paradigmatic example of a good inductive argument, the ravens argument:

The first raven I have observed is black.  
The second raven I have observed is black.  
The third raven I have observed is black.  
[Insert here 96 more cases.]  
The hundredth raven I have observed is black.  
Therefore,  
All ravens are black.

We can compare this paradigmatically reasonable inductive argument with the argument for no fundamental level and see why the latter is less compelling.

First, the argument considered against there being a fundamental level to reality is an induction on only four cases. But inductive arguments we ordinarily take to be compelling tend to involve more than three or four premises. If someone had only seen four ravens, this wouldn't be a good basis for inferring that all whatsoever are black. Second, Sider notes, the inductive argument against a fundamental level moves from a finite group of observations to an infinite conclusion. This is also different from the paradigmatically good inductive inferences we are used to in science and everyday life, where we are instead inferring universal claims about a finite

class of objects. It is questionable whether, if there really were an infinite number of ravens, the inductive argument above would be compelling. Third, Sider notes that in the argument against a fundamental level, the conclusion has a form that is quite different from its premises. He notes the premises of the argument are something like:

The fundamental features were thought to be features of molecules.  
Molecules were rejected as fundamental; the fundamental features were thought to be features of atoms.  
Atoms were rejected as fundamental; the fundamental features were thought to be features of protons, neutrons, and electrons.  
Protons, neutrons, and electrons were rejected as fundamental; the fundamental features were thought to be features of quarks, leptons, and gauge bosons.

These premises would seem to support an inductive conclusion like:

Therefore, anything physicists claim to be fundamental will later be rejected; the fundamental features will be some other class of entities.

But, Sider notes, this is very different from the claim that there is no fundamental level and that reality goes on “forever.” Actually, the conclusion the premises lead to is completely compatible with the existence of a fundamental level, just one the physics community will be forever ignorant of.

Now even if there aren't good arguments in support of a fundamental level, there are nonetheless intuitions in its favor that are difficult to shake. Schaffer once noted that if there were no fundamental level, “reality would be infinitely deferred, never achieved” (2009, p. 376). Many have read Schaffer here as getting at a deep intuition that suggests the concept of infinite descent or “no fundamental level” is incoherent.

However, there is reason to think that the intuition is mistaken. For the idea that reality is deferred until the absolutely fundamental level seems to rely on the assumption that it is only the absolutely fundamental entities that exist, or that exist in some most complete sense.<sup>26</sup> However this is not what those who discuss fundamentality usually believe. Most metaphysicians believe that the entities that are derivative or ontologically dependent on the fundamental are just as real as the fundamental entities. They are generally Quineans (see Chapter 1) at least in the sense that they think there is only one way to exist. And since the derivative entities also exist, they exist in just the same way (and “to the same extent”) as the fundamental entities. This means, whether there is infinite descent or not, reality is not deferred from one level to the next. Each level's entities are just as real as any other's.

In conclusion, there seems to be neither a good argument nor a reliable intuition in support of or against the idea that there must be entities

that are absolutely fundamental. This doesn't mean that it cannot be reasonable for philosophers or scientists to work with the assumption of a fundamental level for the purposes of simplifying a discussion or motivating research. But metaphysicians will have to work harder if they want to show that there is a good reason to maintain such an assumption is ultimately correct.

### **Suggestions for Further Reading**

Two excellent overviews of the mind-body problem (though both precede the literature on grounding) are Jaegwon Kim's *Philosophy of Mind* and Paul Churchland's *Matter and Consciousness*. Kevin Morris's *Physicalism Deconstructed* does an excellent job of updating the discussion to include an assessment of grounding physicalism. A classic defense of property dualism is David Chalmers's *The Conscious Mind*.

Early articles on supervenience may be found in Jaegwon Kim's first volume of collected essays, *Supervenience and Mind*. A compelling statement of supervenience physicalism and how to extend it beyond the mind-body problem is given in Frank Jackson's book *From Metaphysics to Ethics*. A particularly clear statement of the argument that supervenience physicalism does not rule out substance dualism is given in Jessica Wilson's paper "Supervenience-based Formulations of Physicalism."

The four initial papers most responsible for the reframing of metaphysical debates into debates about grounding are Jonathan Schaffer's "On What Grounds What," Kit Fine's "The Question of Realism," Gideon Rosen's "Metaphysical Dependence," and Paul Audi's "Grounding." It should be noted however, that each of these philosophers understands 'grounding' differently. A useful overview is Kelly Trogdon's survey article: "An Introduction to Grounding." Two useful volumes entirely devoted to grounding, its applications, and the debate about the usefulness of the concept are *Metaphysical Grounding*, edited by Fabrice Correia and Benjamin Schnieder and *The Routledge Handbook of Metaphysical Grounding*, edited by Michael Raven. Karen Bennett's *Making Things Up* discusses ontological dependence from a more general perspective.

Tuomas Tahko's entry on fundamentality for the *Stanford Encyclopedia of Philosophy* is an excellent overview of different ways of understanding the notion. Ricki Bliss and Graham Priest's edited collection *Reality and Its Structure* contains many excellent articles on the topic. Jennifer Wang's paper "Fundamentality and Modal Freedom" discusses issues with thinking that fundamentality implies some kind of modal independence. Theodore Sider's *Writing the Book of the World* is a challenging but important book that argues for a particular view about fundamentality and contrasts it with other approaches.

On whether there is a fundamental level, in addition to the discussion cited from Sider's 2011 book, see Jonathan Schaffer's "Is There a Fundamental Level?" and Ross Cameron's "Turtles All the Way Down." Bennett's *Making Things Up* also has a nice discussion of why arguments for an absolutely fundamental level are difficult to come by.

## Notes

- 1 Descartes famously argued for this position in his *Meditations on First Philosophy* (especially Meditations II and VI) from 1641.
- 2 George Berkeley's *Three Dialogues Between Hylas and Philonous* from 1713 is a fascinating defense of idealism over rival metaphysics.
- 3 See the survey results of philosophers presented at <https://philpapers.org/surveys/results.pl>. The rise of phenomenism is related to interest in a position called Russellian monism. For a recent discussion, see Goff (2017).
- 4 One might reasonably ask: how can mental properties be less fundamental than physical properties if they are identical to them? This is a good question. A reductive physicalist might say in response that there is a special class of physical properties which are the fundamental ones, the properties of electrons and other basic entities. Mental properties are identical to less fundamental physical properties, the properties of complex structures of these more basic physical entities.
- 5 This kind of reasoning was later defended by some identity theorists (e.g., Hill 1991), and questioned by others (e.g., Kim 2005).
- 6 Other influential accounts may be found in Lewis (1972), Wilson (1999), Gillett (2002), Melnyk (2003), and Shoemaker (2007). Melnyk's 2003 book *A Physicalist Manifesto* is a comprehensive defense of realization physicalism.
- 7 Frank Jackson's book *From Metaphysics to Ethics* does an excellent job of describing and defending such a comprehensive form of supervenience physicalism.
- 8 This phrase comes from (and plays a big dialectical role in) David Armstrong's 1997 book *A World of States of Affairs*.
- 9 G.E. Moore is one philosopher who endorsed such a view (Moore 1903).
- 10 Though Wilson (2014), Barnes (2018), and Bliss (2018), among others, debate this point.
- 11 This isn't to say everyone agrees. There are philosophers of mind who continue to insist that supervenience is adequate to formulating a physicalist position. See, for example, Howell (2009).
- 12 Schaffer actually makes the stronger claim that *all* of the important issues in metaphysics are issues about grounding. This is much more controversial.
- 13 Note there is actually more than a terminological disagreement between Wilson and Bennett here. Although they agree that there are a variety of metaphysical relations at issue, they disagree about these relations' formal features. For Wilson counts identity as one example of a grounding relation. But of course, as we discussed in Chapter 3, identity is a symmetric relation. Bennett argues that building relations cannot be symmetric.
- 14 By 'first-order projects,' I mean projects about what is grounded in what, not about what grounding itself is.
- 15 For example, the physicalist David Lewis is explicit in his work that the fundamental physical properties are freely recombinable in this way with each other (e.g., Lewis 1986).
- 16 See Jessica Wilson (2010) for discussion.
- 17 There are exceptions. In Chapter 2, we discussed nominalists who only believe in objects, no properties. There we also discussed Plato, whose fundamental ontology contained only the Forms (properties). But many metaphysicians hold at

least a two-category ontology which includes as fundamental both objects and properties.

- 18 To cite just one issue (for more see Ladyman 2017), in quantum field theory, subatomic particles like electrons and protons are typically not understood as localized to one tiny region but rather as fields spread out over all of space. But as we saw in Chapter 3, mereological relationships presuppose a decomposition of things into parts that do not overlap in space.
- 19 An example that could illustrate this case is the relationship between particles and their sums brought on by quantum entanglement. See, for example, Ismael and Schaffer (2020) and Ney (2021a).
- 20 See especially Schaffer (2010). Elizabeth Barnes (2012) draws some important and related distinctions.
- 21 Note that this priority monism is different from another position that Schaffer calls 'existence monism.' Schaffer's priority monist believes there is a one that is more fundamental than its parts, but that the parts still exist. Existence monists (like Spinoza or Parmenides in the history of philosophy) believe there is *only* the one.
- 22 See Schaffer (2010) and Ismael and Schaffer (2020) for the argument, and Ney (2010) and Calosi (2018) for rebuttals.
- 23 A classic source on reduction in the philosophy of science is Nagel (1961).
- 24 Sider (2011) criticizes truthmaking in his chapter on "Rivals." See also Fine (2012).
- 25 Schaffer has argued that principles like Ockham's Razor only recommend parsimony with respect to fundamental entities (Schaffer 2015). So the fact that the grounding theorist has nonfundamental entities in their ontology doesn't detract from the view's parsimony in any sense that matters.
- 26 Bennett makes this point in her book (2017, pp. 120-122).

# 8 Time

## Learning Points

- Introduces and evaluates several rival ontologies of time
- Considers the consequences relativity theory has for these ontologies
- Introduces the distinction between the A- and B-theories of time
- Evaluates the prospects for the logical possibility of time travel

## 8.1 Time's Passage

A basic and perfectly ordinary fact about time would seem to be that it passes. But what do we mean when we say this?

Consider the most important and meaningful events in your life up to now: your birth, your first day of school, a moment when you overcame a great challenge, the time you first bonded with your closest friend, perhaps the loss of a loved one. Whichever these important events are, they are all events lying in your past. At one time they were present, at one time you were living them, you were a part of them. But now they are over. Time has passed. Time and you yourself have in a sense moved beyond these events. It is important to you that they happened; these are events that have shaped you and who you are. But in some sense, these events are no longer happening. They are no longer real. But they *did* happen. They *were* real.

Now consider what lies ahead, those events in your future. We can be certain the future will include many mundane events like the sun's rising and your next meal. There are also some more significant events we can be certain your future will include, including your death. Still, there is a lot of uncertainty about the future, more so than there is about the past. Can you be sure what you will be doing five years from now? Whether you will

be a success? Can you be sure who the important people in your life will be ten years from now? Where you will be living? These are events most of us don't think we can be certain about, because we haven't lived them yet. They haven't yet happened. They are events we can only plan for or hope to be certain ways, because they are events that only occur in the future. Time has not yet caught up to them. Unlike events in the past and present, it is very natural to think that there is not yet any fact about what the future will be like. We may have reasonable beliefs, even certainty, about some events in the future (like the sun's rising), but since the future has not yet come to pass, there aren't yet any determinate facts about this future. Philosophers refer to this as the **openness of the future**: the future is open in a way the past and present are not.

**Openness of the future:** the view that there are not any determinate facts about the future

This is part of what we ordinarily have in mind when we think that time passes. There are events in the past, those that have already happened. These are fixed; the past is closed. Then there is what is happening now. And finally, there is what has not yet happened – the future is open. Time's passage consists in new events coming to be and becoming present, and those events that are present slipping into the past. (And those that are already past slipping even further into the past.) We ordinarily think time's passage is objective or absolute. It doesn't depend on your own personal perspective. Time doesn't pass differently for different people. And it is not subject to anyone's control. Indeed, we ordinarily think time would pass in just the same way as it actually does even if humans never existed.<sup>1</sup> And once an event has happened, it belongs in the objective past.

### EXERCISE 8.1

#### The Ordinary View of Time's Passage

List four parts of the ordinary, what we might call the common-sense, view of time's passage.

**A-theory of time:** the view that tensed facts are irreducible to and fail to be grounded in tenseless facts

**B-theory of time:** the view that tensed facts are reducible to or grounded in tenseless facts

This view, that time's passage is absolute, and that there is an objective distinction between which events are past, which are presently happening, and which lie only in the future, is what philosophers call the **A-theory of time**. It will become clear why it is called the A-theory later, once we have introduced some more distinctions.

The A-theory of time is most commonly contrasted with the **B-theory of time**, which takes time's passage to be a subjective phenomenon, one that is tied to individual perspectives. According to the B-theorist, although it may be true for us now that the U.S. Civil War lies more than a hundred and fifty years in the past, this fact is subjective. There are other

individual perspectives from which this is not the case. From the perspective of the first U.S. President, George Washington, the Civil War lies only in the future. From the perspective of the sixteenth U.S. President, Abraham Lincoln, the Civil War lies in the present. According to the B-theory, no one of these perspectives is more correct than any other. There is no absolute, perspective-free fact about what time is now, any more than is an absolute perspective-free fact about which place is here. Just like you and I can have different, but equally correct views from our perspectives about which place is “here,” people at different times can have different, but equally correct views from their perspectives about which time is “now.” This is a revisionary view about the passage of time, since it makes it a subjective, rather than an objective, phenomenon. B-theorists cannot say, like the A-theorist can, that the U.S. Civil War is objectively in the past, and that anyone who thinks it is in the present or future is wrong.

The B-theory of time is typically combined with an ontological view about time. This is called **eternalism**. Eternalism is the view that past, present, and future objects and events are equally real. For the eternalist, reality is not three-dimensional, but four-dimensional. In addition to there being all of these many places in our universe (California, Paris, Hong Kong, the planet Venus, the Andromeda Galaxy), there are also many times. Reality, according to the eternalist, is a four-dimensional spatiotemporal block. Just as we ordinarily think objects and events at other places are real (the book or computer in front of you is no more real than the Eiffel Tower), the eternalist thinks objects and events at other times are real (you are no more real than Abraham Lincoln).

The combination of eternalism with the B-theory of time also has a name: it is typically called **the block universe view**. According to this view, reality is a four-dimensional spatiotemporal block in which all objects and all times are equally real, *and* there is no objective fact about which of these objects and events are present. That is, there is no objective fact about where in the block the “now” is.

It is possible to combine eternalism with the other view about time’s passage, the A-theory, although this combination of theories is much less popular than the block universe view. The combination of eternalism with the A-theory of time is what is known as **the moving spotlight view**. Here again, past, present, and future objects are all equally real, but since this is an A-theory, there is an objective additional fact about where the “now” is. (See Text Box 8.1 for more on the moving spotlight view and other ways of combining the A-theory with ontologies of time.)

More commonly, when philosophers endorse an A-theory about time’s passage, they hold a distinct ontology of time. Most often, they are **presentists** rather than eternalists. Presentism is the view that the only objects and events that exist are those that are present. For the presentist, it is straightforward why there is an objective fact about which time

**Eternalism:** the view that past, present, and future objects and events are equally real

**Block universe view:** the combination of eternalism and the B-theory of time

**Moving spotlight view:** the combination of eternalism and the A-theory of time

**Presentism:** the view that only present objects and events are real



is present – it is because the present constitutes everything that exists. It isn't that George Washington and Abraham Lincoln are wrong to think the U.S. Civil War is in the future or the present. For the presentist, George Washington and Abraham Lincoln don't exist. They *did* exist, and when they did exist, they had generally correct views about which events were past and present, and which events, if they were to come to pass, would lie only in the future.

Thus for the presentist, time's passage has ontological implications. Those people in the past who have died and those objects that have been completely destroyed do not exist. You exist, but your great great grandmother does not. The European Union exists, but the Ottoman Empire does not. Similarly, no wholly future people or objects exist. You may have hopes, make plans, or suffer fears about which objects or events may exist in the future – future offspring, wars, jobs. But it is part of the presentist's way of thinking about the future that these things do not exist. They are not real.<sup>2</sup>

### TEXT BOX 8.1

#### Temporal Metaphysics

There are four metaphysics of time that are most commonly discussed:

1. **Presentism**

Presentism is an A-theory metaphysics that combines the view that the passage of time is objective and absolute with an ontology according to which only present objects and events exist.

2. **The growing block theory**

The growing block theory is an A-theory metaphysics that combines the view that the passage of time is objective and absolute with an ontology according to which both past and present objects and events exist, but future objects and events do not. Reality is a four-dimensional block that grows as time passes. According to this view, the present is metaphysically privileged in being the cusp or edge of reality.

3. **The moving spotlight view**

The moving spotlight view is the combination of the A-theory with eternalism, the view that past, present, and

**Growing block theory:**  
the view that past and present objects and events are real, but future ones are not

future objects and events are equally real. As an A-theory metaphysics, it includes the view that the passage of time is objective and absolute. This is accommodated by stipulating that present objects and events have some metaphysically distinctive feature (existing or occurring now) that past and future objects and events lack. But this metaphysically distinctive feature is not the present entities' existence, since according to the moving spotlight theorist, past and future entities exist as well.

#### **4. The block universe view**

The block universe view is the combination of the B-theory with eternalism, the view that past, present, and future objects and events are equally real. As a B-theory metaphysics, it includes the view that the passage of time is subjective and relative to an individual's perspective. According to the block universe theorist, you and people in the past equally exist, but each have distinct perspectives on which time is now, none of which is more correct than any other.

In this chapter, we will primarily focus on arguments for and against the two most dominant positions, presentism (which is widely taken to be the common sense view) and eternalism (the revisionary one).

### **EXERCISE 8.2**

#### **Four Metaphysics of Time**

Of the four metaphysics of time outlined in Text Box 8.1, according to which is it true that:

1. Socrates exists.
2. The events of the year 2200 AD exist.
3. Currently existing objects possess some metaphysical distinctiveness that past objects lack.
4. Some people are correct in believing it is currently the twentieth century.
5. Nobody is correct in believing it is currently the twentieth century.
6. There exist some people who believe it is currently the twentieth century, but they are wrong about that.

Presentism is often thought to be a more precise statement of our ordinary conception of time and its passage. It is the common sense view about time. Our focus in the next two sections will be to investigate which if any arguments would make one think that we should reject this common sense view about time and replace it with eternalism and the block universe view. There are two main arguments in favor of the block universe view that metaphysicians have taken seriously. The first argument, the Argument from Special Relativity, challenges presentism by claiming it is incompatible with one of our best, empirically supported scientific theories. The second, the Truthmaker Objection, challenges presentism by claiming that presentists have no convincing way to provide a metaphysical grounding for truths about the past and future. We will begin by examining the Argument from Special Relativity.

## 8.2 Background on Special Relativity

The Argument from Special Relativity purports to show that the 1905 theory of relativity of Albert Einstein (1879-1955) provides a decisive reason to endorse the block universe view over presentism.<sup>3</sup> To understand the argument, we will need to get a basic understanding of Einstein's theory.

The theory, as presented by Einstein,<sup>4</sup> has two central postulates:

1. The laws of physics are the same in all inertial reference frames.
2. The speed of light is the same for all observers.

### **Inertial reference**

**frames:** systems in which objects are moving at constant velocity

By **inertial reference frames**, we just mean systems in which objects are moving at constant velocity, in other words, systems in which the objects are not accelerating. It is Einstein's general theory of relativity that discusses accelerating objects (including objects that are accelerating under the influence of gravity), but for this argument we will just focus on the special theory of relativity that only deals with objects moving at constant velocity. To illustrate what the first postulate is saying, imagine we have two systems moving at constant velocity (two inertial systems): one is a person who is at rest (say, sitting on a park bench) and another is a person running at a constant velocity of 10 kilometers per hour relative to the seated person. The first postulate says that it isn't the case that the laws of physics are different for the person on the bench and the runner. The same laws of physics apply to both of them. It's a very simple postulate.

The second postulate is more interesting. To see what is interesting about it, suppose a second runner is also moving at 10 kilometers an hour relative to the person on the bench, but in the opposite direction. (See Figure 8.1.)

Now consider the questions:

1. What is the velocity of Runner 1 relative to the person on the bench?
2. What is the velocity of Runner 1 relative to Runner 2?



Figure 8.1 Inertial Frames

The answer to Question 1 is easy. Relative to the person on the bench, the velocity of Runner 1 is 10 kilometers per hour. But the answer to Question 2 is different, since Runner 2 is running towards Runner 1. To figure out the answer to Question 2, notice that if Runners 1 and 2 had initially started very far apart, then we know that after an hour, at the paces they are running, they would be 20 kilometers closer to each other than they were initially. This is because after an hour, Runner 1 would move 10 kilometers in Runner 2's direction and Runner 2 would move 10 kilometers in Runner 1's direction, and so they would get a total of 20 kilometers closer in that hour. So, the answer to Question 2 is that the velocity of Runner 1 relative to Runner 2 is 20 kilometers per hour. The speed of a runner is relative to one's reference frame.

Now we can return to Einstein's second postulate. What the second postulate says effectively is that light is not like runners (or anything else!). Einstein's bold conjecture was that light moves at the same speed relative to all observers. So, whether you are sitting on a bench, running towards a light source or running away from it, the velocity of light relative to your perspective will be the same. This is surprising, and departs from what was assumed by the earlier Newtonian physics that preceded Einstein's theory. Nonetheless, the conjecture has turned out to be confirmed repeatedly by experiment.

We will need these two postulates of special relativity in order to derive what is the key premise in the argument against presentism. The key premise is that there are no absolute facts about whether two events are simultaneous. Facts about whether two events are simultaneous or whether the first happens before the second or vice versa are all, according to the special theory of relativity, relative to one's inertial frame.

To see how the two postulates of special relativity entail this premise, we will make use of a famous thought experiment. This is a thought experiment that Einstein liked to use to present his theory and its surprising consequences. In this thought experiment, we imagine there are two observers. Again, suppose the first one is seated on a bench. Call him 'Bench.' But this time the second one is moving on a train (again at

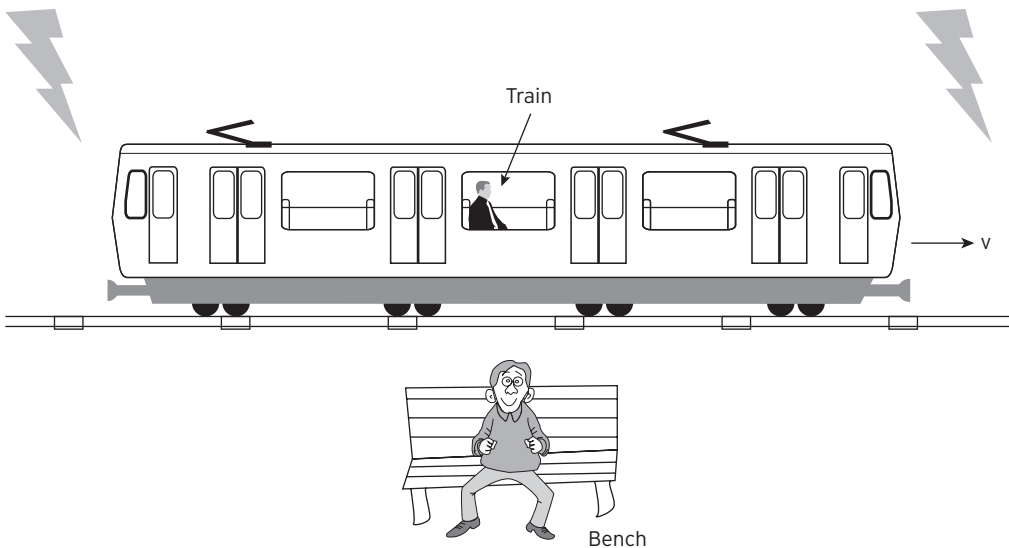


Figure 8.2 Einstein's Thought Experiment

constant velocity). Call him 'Train.' Train is moving at some significant velocity relative to Bench. Einstein wants us to imagine that there are two lightning strikes. They flash at spatial locations that are the same distance from Bench, and Bench sees the flashes at the same time. But from Train's perspective, one flash is behind him and the other is in the direction he is moving towards. Train sees one strike before he sees the other.

The question we now want to ask is whether the lightning strikes are simultaneous (occur at the same time) or not. Let's start with what Train should say. Train's verdict about the simultaneity of the lightning strikes depends on whether he assumes Einstein's second postulate and the theory of special relativity, or whether he assumes that the relative speed of light varies depending on the motion of the observer, as the earlier Newtonian theory would say.

Let's start by assuming Newtonian physics. Using Newtonian physics, Train will reason that because he is moving away from one strike, the light from it will take longer to reach him (it moves at a slower relative speed). Since he is moving toward the other strike, the light from it will reach him sooner (so it is moving at a faster relative speed). So, even though we may suppose Train sees the light from the one strike before he sees the light from the other, this doesn't rule out his inferring that the two strikes occurred at the same time. He can calculate using Newtonian physics and his observation of the spatial locations of the flashes, that the lightning strikes occurred at the same time. He sees them at different times because the light traveling from the strikes have different velocities, relative to his frame of reference.

Things change, however, when we introduce Einstein's second postulate. Now Train is not entitled to assume that the light from the two strikes moves at different velocities, relative to his frame of reference. Instead, he must assume that even though he is moving away from one of the strikes and toward the other, light always moves at the same speed relative to him. And so from this fact, combined with his observation of the spatial locations of the flashes and the fact that he observed one after the other, he is forced to conclude that the one occurred later than the other. If Train assumes the special theory of relativity, he will conclude that the lightning strikes were not simultaneous.

Now let us consider what Bench should say. Bench sees that the strikes occur at different locations in the sky, but each is the same distance away from him. Whether Bench assumes Newtonian or relativistic physics, he will infer that the light moves at the same speed, relative to him. So, the disagreement between these two theories doesn't come into play. Since the light reaches Bench at the same time, he will reason that the two lightning strikes were simultaneous.

Let's summarize these results:

Using Newtonian physics:

Bench infers that the two lightning strikes were simultaneous.

Train infers that the two lightning strikes were simultaneous.

Using relativistic physics:

Bench infers that the two lightning strikes were simultaneous.

Train infers that the two lightning strikes were not simultaneous.

This means that according to the theory of relativity, observers in different inertial frames (e.g., Bench and Train) will use the same laws of physics to derive different facts about the simultaneity of events. The first postulate of special relativity says that the laws of physics are the same in all inertial frames, so this means, if relativity is correct, then the facts about simultaneity are different in different inertial frames. There are no absolute facts about the simultaneity of events (whether they happen at the same time or not). We now have what we need to present the argument from relativity against presentism (and in defense of the block universe view).

## TEXT BOX 8.2

### The Metaphysics of Special Relativity

According to special relativity: fundamentally, reality doesn't consist of a collection of objects spread out in a three-dimensional space that persists from one instant to the next through time. Instead, reality is a collection of events taking place in one unified four-dimensional

manifold of spacetime. Objects may be thought of as lines through the manifold, or sequences of events (more on this in Chapter 9). The spacetime of special relativity is what is called 'Minkowski space-time,' after the mathematician Hermann Minkowski (1864-1909), who suggested the idea to Einstein. Minkowski spacetime has the following important features:

- It is a four-dimensional manifold. In other words, in order to specify a location in Minkowski spacetime, one has to give four numbers:  $(x, y, z, t)$ .
- There is no preferred, objective partition of this manifold into slices of time. (Different observers moving at different relative velocities will each slice spacetime up into spaces and times in different, equally correct ways.)
- There are objective facts about the distances between spacetime points, but no objective facts about spatial distances and temporal durations. (Facts about spatial distances and temporal durations are always relative to a particular observer's way of carving up spacetime.)

To illustrate this third point, we may return to our two lightning strikes. Each strike takes place at a distinct location in Minkowski spacetime. What the third point implies is that there is an objective fact about the spatiotemporal distance between the strikes, but no objective fact about their temporal separation.

One way to think about this is that different observers may slice up the one spacetime manifold that exists into different simultaneity or time slices. For example, when the spacetime manifold is sliced up in the way suggested in Figure 8.4, there appears to be a fact that one lightning strike happens before the other.

One lies on an earlier slice of spacetime and so they are not simultaneous. What special relativity teaches is that the slicing depicted in Figure 8.4 is not part of the objective structure of our world. Just as one observer may slice things up in this one way, another observer moving at a different velocity will slice things up in another way. For this

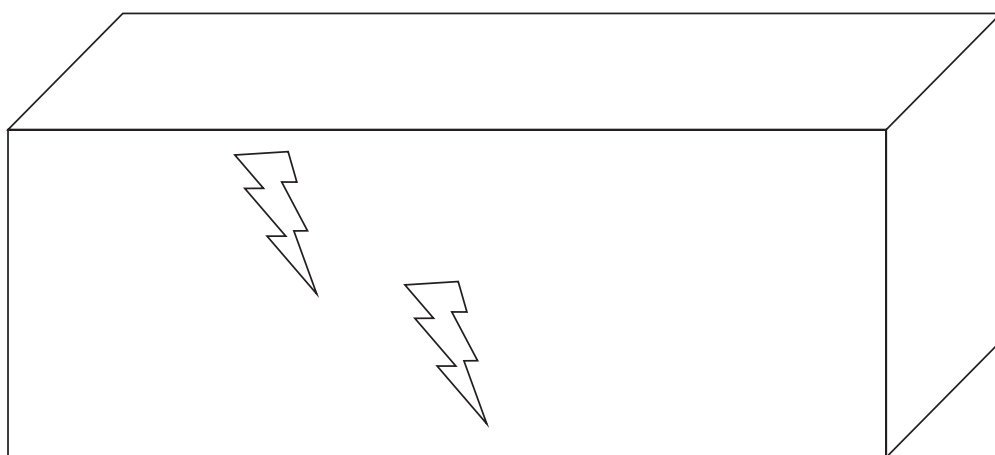


Figure 8.3 Minkowski Spacetime

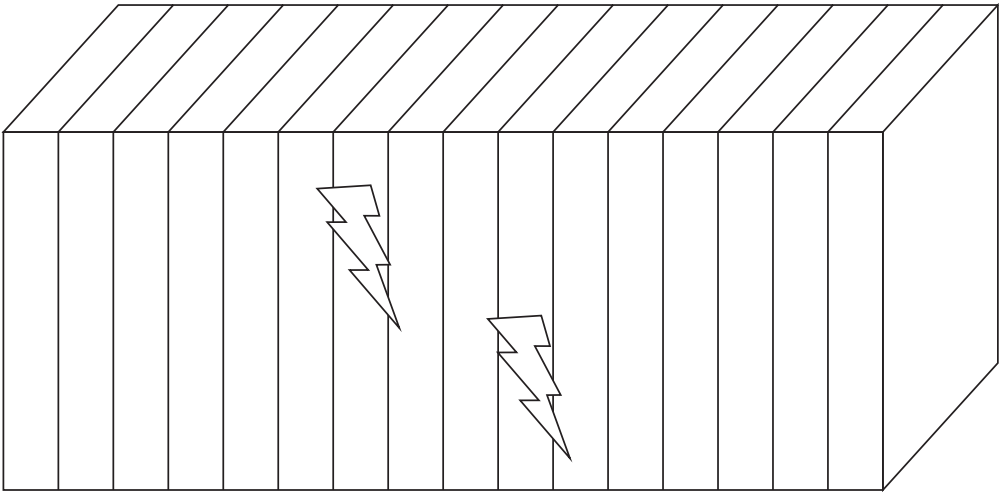


Figure 8.4 Spacetime Containing Absolute Facts about Simultaneity

reason, the objective structure of spacetime given by modern physics is better captured by Figure 8.3 than Figure 8.4, and this has consequences for which ontology of time we ought to accept. If we want to draw ontological lessons from special relativity, then we should not accept an ontology that imposes an objective distinction between past, present, and future.

### 8.3 The Argument from Special Relativity Against Presentism

Here is the argument. It has the form of a *reductio ab absurdum*. (See Section 3.2.)

1. Assume that presentism is true.
2. Then, the only events that exist are those that are present.
3. The events that are present are those that simultaneous with what is happening here and now.
4. There are no absolute facts about simultaneity; rather, whether two events are simultaneous is always relative to one's frame of reference.
5. Therefore, which events are simultaneous with what is happening here and now is always relative to one's frame of reference.
6. Therefore, which events are present is always relative to one's frame of reference.
7. The facts about what exists are not relative to one's frame of reference.
8. Therefore, presentism is false.



The three most interesting premises in this argument are the independent premises (3), (4), and (7). These are what someone who wants to use this argument to argue against presentism (and in defense of the block universe view) will need to motivate.

### TEXT BOX 8.3

#### Generalizing the Argument from Special Relativity

As presented earlier, the argument from special relativity is an argument against presentism. However, this argument is easily generalized to challenge any temporal ontology according to which the present is metaphysically privileged, that is, any A-theory metaphysics. The presentist argues that the present is metaphysically privileged in comprising all that is real. The growing block theorist believes that the present is metaphysically privileged in being the cusp or ultimate edge of reality. How the present is metaphysically privileged according to the moving spotlight theory is less clear; but the metaphysical distinctiveness of the present is still an essential aspect of that view as well.

The general form of argument in favor of the block universe view over any of these A-theory metaphysics may then be stated in the following way:

1. Assume that some A-theory metaphysics is true.
2. Then, there is something metaphysically distinctive about the present.
3. The events that are present are those that simultaneous with what is happening here and now.
4. There are no absolute facts about simultaneity; rather, whether two events are simultaneous is always relative to one's frame of reference.
5. Therefore, which events are simultaneous with what is happening here and now is always relative to one's frame of reference.
6. Therefore, what is metaphysically distinctive is relative to one's frame of reference.
7. What is metaphysically distinctive is not relative to one's frame of reference.
8. Therefore, no A-theory metaphysics is correct. Thus, one should endorse the block universe view of time.

Of the three key premises, (3) and (4) are the same as those discussed in the main text. (7) is a more general version of the premise in the argument against presentism. The defender of this argument will say: not only are the facts about what exists objective, and not relative to anyone's perspective, but so are more generally the facts about what is metaphysically privileged. This claim will be plausible for most metaphysicians, as metaphysical privilege is typically understood by definition to imply objectivity. This argument leaves the block universe view as the only standing metaphysics of time, since it is the only theory that does not make reality depend upon frame-dependent or perspectival facts about what is past, present, or future.

Premise (4) was the topic of the previous section. The thought experiment with the two lightning strikes was introduced by Einstein specifically with the aim of showing how his special theory of relativity implies the relativity of facts about simultaneity.

Thus, it just remains for us to discuss the motivations for premises (3) and (7). Premise (3) is supposed to be an analytic truth about what it means to say an event is “present.” Here is the idea. If a person considers which events are the present ones, they obviously must include all of those events that are happening to them “right here,” at the location they currently occupy, and “right now,” at the current time. But assuming the person is not a solipsist, these are clearly not the only events they will count as happening in the present, for there are others occurring at other spatial locations. Which other events are the present ones beyond those happening in the here and now? The natural answer is all of the events that are happening at the same time as (simultaneous with) those that are happening in the here and now. This is just what premise (3) says.

Finally, let us discuss premise (7), that existence or reality is an objective matter and so does not depend on one’s perspective. How could it be that an entity is real relative to one person, and unreal according to another? If the entity is real relative to one person, then it is real. The other may not believe in it, but if the object is out there to be believed in, then it exists. Following the standard Quinean view about existence developed in Chapter 1, it doesn’t make sense to say that objects can exist only relatively. Presentists and eternalists usually agree with this premise.

Since the argument against presentism from special relativity has three substantive premises, there are three natural ways that a presentist could respond to this argument, claiming it is not sound. They could deny premise (3), claiming that this is not the right way to understand what ‘present’ means. They could deny premise (4), arguing either that special relativity is wrong, or that it does not have the consequence about simultaneity that Einstein thought it did. Or they could deny premise (7), and argue that facts about existence are not, as a matter of fact, absolute.

The most common way for presentists to respond to this argument is to take on premise (4) either by rejecting special relativity or claiming that special relativity does not have the consequences most physicists think it has (see e.g., Craig 2001, Markosian 2004, Zimmerman 2011). Metaphysicians may think that it is so obvious that present events and objects exist in a way that past and future ones do not, that no scientific theory could ever give us good reason to undermine this claim. And so either special relativity is wrong, or it can’t have the consequences this argument suggests. Premises (3) and (7) are analytic and rely on straightforward and pretty much unanimously agreed views about what is meant by ‘present’ and ‘exists.’<sup>5</sup>

**EXERCISE 8.3****The Argument from Special Relativity**

If you were a presentist, how would you respond to the argument from Special Relativity? How would you show the argument is not sound? Be sure to single out an independent premise or inference step in the argument.

**8.4 The A-Theory and the B-Theory**

Those who reject presentism are typically eternalists. When one views the universe as the eternalist does, one believes that future events are as real as present and past ones, and so in a certain genuine sense, the future is already set. The future is set in the sense that there are facts about which events take place in the future.<sup>6</sup> To say this without getting into trouble, the eternalist will often find it helpful to explicitly distinguish two senses of 'is' and 'exists.'

To see this, consider the following sentences:

- (1) Abraham Lincoln, unlike Sherlock Holmes, exists. Abraham Lincoln, unlike Sherlock Holmes, is a real person.
- (2) Elon Musk, unlike Abraham Lincoln, exists. Abraham Lincoln, unlike Elon Musk, isn't real anymore.

At first glance, (1) and (2) look to both be true, but they also seem to contradict one another. One says that Abraham Lincoln exists; the other says he doesn't. Different philosophers will address the apparent tension between (1) and (2) in different ways. The presentist will resolve the tension by denying that (1) is true. According to the presentist, past objects fail to exist. Abraham Lincoln is no more real than Sherlock Holmes. The eternalist, on the other hand, will want to say that both (1) and (2) are true, and deny that there is a contradiction between them. In order to do so, the eternalist, will appeal to two senses of 'exist' and 'is.'<sup>7</sup>

The sense of 'is' and 'exists' used in (1) is the tenseless or eternal sense. Something exists in the tenseless or eternal sense of the word just in case it exists at some time or other, or, if there is a realm of existence outside of the spatiotemporal manifold, then it exists in that realm.<sup>8</sup> Sentence (2), on the other hand, uses 'exists' in a tensed sense. Here, 'exists' means *exists now*. Since the eternalist believes there are objects and events that exist in the tenseless sense that do not exist now, she is free to endorse both (1) and (2) without contradiction. The presentist, who thinks that existence is limited to what exists now, will reject this

distinction between two senses of 'is' or 'exists.' For her 'exists' means the same as 'exists now.'

One might then view the eternalist's position this way. First, there is existence in the tenseless, eternal sense. There is what just plain exists – everything in the four-dimensional block. But there are also facts about what exists in the block at certain points, facts about what exists now or earlier or later than now. The eternalist who endorses special relativity will note that facts about what exists in the past, present, and future are all relative to one's frame of reference. This means the fact about what exists now and so what exists in a tensed sense of 'exists' is subjective or perspectival. But the fundamental facts about what exists are not subjective. These are the facts about what exists in the tenseless, objective sense of the word. And that is what matters for ontological purposes.

Many eternalists also (importantly) argue that tensed facts about what is past, present, and future are reducible to, grounded in, or made true by more fundamental tenseless facts.<sup>9</sup> This marks an important disagreement between eternalists and presentists. To see this, we can now finally give a more official account of the distinction between the A- and B-theories of time.

This distinction traces back to a 1908 paper by the English philosopher J.M.E. McTaggart (1866-1925).<sup>10</sup> McTaggart distinguished between the A-series and B-series of time. Both are series in which events are ordered according to their positions in time.

First, the A-series. The **A-series** orders events in terms of their being past, present, and future (their tensed, **A-features**). We may picture the A-series as a long timeline with those events that are farthest in the past all of the way at one end, and those events farthest in the future all of the way at the other end. (If time is infinite, then we may imagine the A-series as an infinite line with no beginning or end.) Here are some statements we may make assigning objects and events places in the A-series:

- Washington is the capital of the United States *now*.
- In the *present* day, people check the time using their phones.
- We *will* go to the shore next summer.
- The Sun *will* begin to die approximately 5 billion years into the *future*.

Events' locations in the A-series change as time passes. What is in the future now, will be in the present eventually. And what takes place now, will eventually be in the past. We may call those facts that attribute a location in the A-series to objects and events (such as those listed above): A-facts.

The **B-series** is another temporal ordering, but here events are ordered in terms of their dates and times, as well as their relations of being earlier, later than, or simultaneous with other events. They are ordered according

**A-series:** a temporal ordering of events as being past, present, and future

**A-features:** tensed features, features that change as time passes

**B-series:** a temporal ordering of events as being earlier than, later than, or simultaneous with each other

**B-features:** tenseless but temporal features that do not change as time passes

to their **B-features:** tenseless features indicating relative temporal relations to one another. For example, one might make the following statements relating certain events to a particular year, 2023:

- Washington is the capital of the United States in the year 2023.
- In 2023, people check the time using their phones.
- We go to the shore in the summer a year after 2023.
- The Sun begins to die approximately 5 *billion years after 2023*.

Although events' locations in the A-series change, events' locations in the B-series do not. As time passes, it remains a fact that Washington is the capital of the United States in the year 2022. Even in the year 3022, it will remain the case that Washington is the capital of the United States in 2022. Note in all of the B-statements, we are using our verbs - 'is,' 'check,' 'go,' and so on - tenselessly. To parse them, it may be helpful to imagine an observer who stands outside of time and is able to view the entire B-series at once. This observer will be able to make statements about when certain events occur, and what are these events' temporal relations to one another. Those facts that attribute locations in the B-series to objects and events are the B-facts. Unlike A-facts, B-facts do not change. And, as we have just seen, these are tenseless facts in the sense that they use tenseless verbs, and are true eternally from an objective perspective. Nonetheless although B-facts are tenseless, they are facts about time since facts about the dates and times of events and their occurring earlier or later than each other are facts about these events' locations in time.

Except for those nihilists who deny the reality of time altogether,<sup>11</sup> metaphysicians agree that there are these two kinds of facts about time: A-facts and B-facts. The big debate concerns which of these two kinds of facts are more fundamental, the A-facts or the B-facts.

The B-theory of time states that all A-facts (facts about what is past, present, and future) are made true by or grounded in the B-facts. Another way to put the B-theorist's position is this: the (tenseless) B-facts are more fundamental than the (tensed) A-facts. Let's consider a couple of examples to illustrate the B-theorist's position. To formulate these examples, we will make use of the concept of grounding introduced in Chapter 7, but nothing will hang on this choice. Those who prefer to understand relative fundamentality in terms of building relations or truthmaking instead should feel free to reformulate the following into their favored idiom.<sup>12</sup>

Pretend that today's date is May 9, 2023. A B-theorist would say that the fact expressed by the following sentence:

(1A) Washington is the capital of the United States now,

is ultimately grounded in the fact expressed by this sentence:

(1B) Washington is the capital of the United States on May 9, 2023.

A fact about what happens now, an A-fact, is in this way shown to be nothing over and above a fact about what happens on May 9, 2023, a B-fact. Likewise,

(2A) We will go to the shore next summer,

expresses a fact that, according to the B-theorist, is grounded in the tenseless fact expressed by:

(2B) We go to the shore in the summer in the year after May 9, 2023,<sup>13</sup>

assuming again that (2A) was asserted on that date, May 9, 2023.

#### EXERCISE 8.4

##### B-Theoretic Reductions of A-Facts

All of the following sentences express A-facts. Given the time at which you are doing this exercise, provide a B-theoretic reduction of these facts.

- A. Dinosaurs once roamed the Earth.
- B. At one time, Philadelphia was the capital of the United States, but now the capital of the United States is Washington D.C.
- C. In the future, there will be human outposts on Mars.
- D. The bell rang five minutes ago.

The opposing view to the B-theory is the A-theory of time. The simplest way to understand the A-theory is as a rejection of the B-theory. The B-theorist says that B-facts are more fundamental than A-facts. The A-theorist denies this, and says that the B-facts are not more fundamental than the A-facts. According to the A-theorist, the A-facts fail to be grounded in or made true by the B-facts. The facts about the past, present, and future are something over and above tenseless B-facts.

The motivation for the A-theory can be stated simply. A-facts are facts that change. An event that is future will eventually become present. An event that is present will eventually become past. But B-facts never change. If World War I occurs later than the Trojan War, it is always the case that it is later than the Trojan War. If the Los Angeles Olympics occur in July and August of 2028, then they always occur in July and August of 2028. This is characteristic of B-facts. Since A-facts change, but B-facts do not, the A-theorist argues that A-facts are something over and above the B-facts. To

accommodate the genuine, objective sense in which time passes, one needs A-facts, and so there must exist A-facts in addition to the tenseless B-facts.

In his article on the subject, McTaggart endorses what is an even stronger claim. McTaggart (1908) argues not only that facts about past, present, and future aren't grounded in the B-facts, but that indeed the A-facts are *more* fundamental than the B-facts. For McTaggart, if there were no A-facts, no facts about which events are past, present, and future, then there would be no such thing as time – time would be unreal. The reason for this is that time essentially involves change. For time to be real, something must change. But McTaggart thought change is only possible if there is an objective A-series. If there were just the B-facts, facts about the absolute and relative locations of events in the B-series, there wouldn't be any change. McTaggart illustrates this point using a specific example of an event, the death of Queen Anne.

McTaggart asks how, if the fundamental facts about time are facts about the B-series, there could be change. Since an event's place in the B-series is fixed and permanent, events never change their B-features. If the death of Queen Anne occurs in 1714, then it (tenselessly) always occurs in 1714. If the death of Queen Anne occurs before the American Revolution, then it (tenselessly) always occurs before the American Revolution. Events don't change in their nontemporal features either. As he puts it, Queen Anne's death is always a death. The event doesn't exist for a little while, then gain the feature of being a death. The death of Queen Anne is never not a death. If it was a painful death, it was always a painful death. If it occurred in Kensington Palace, then it always occurred in Kensington Palace. McTaggart thus concludes: if there is only a B-series, there can be no change. Change requires the existence of an irreducible A-series. For the only way for events to change is to change in their A-properties. In the case of the death of Queen Anne, this event once was future and then present, but now it is past. At the time McTaggart was writing, it was nearly 200 years in the past. It is now more than 300 years in the past. Because one needs the A-series for there to be genuine change, and, according to McTaggart, there needs to be genuine change if time is to be real, it follows for him that the existence of any temporal series, including the B-series, requires an A-series.

Because many philosophers have followed McTaggart in this reasoning, the B-theory of time is often pejoratively referred to as a static theory of time. Without fundamental facts about time that involve the A-series, it is thought, nothing changes.

This is a legitimate concern about the B-theory. But the B-theorist may respond that this argument has left out of consideration one way in which there can be change in the universe. Yes, events don't ever change according to the B-theory. But events are not the only component of the B-theorist's ontology. The B-theorist may also believe in the existence of

*objects* and say that while events never change their properties, objects do. For at one time (or spacetime point) an object may have a certain property. Then at another time (or spacetime point) that same object may lack that property. To use McTaggart's example again, perhaps the *event* that is Queen Anne's death may not change any of its features. But Queen Anne herself did change. After all, she was alive in 1713 and dead in 1714. Since Queen Anne herself has different properties at different times, there is change after all on the B-theory.

In this way, the B-theorist is able to respond to the worry that their picture is one of a static universe without change. Their response is that since objects in the universe have different features at different locations in the four-dimensional block that is the universe, the view does accommodate change. However, A-theorists have historically been dissatisfied with this response. They may concede that there is a sense in which the B-theorist can accommodate change, but it is not the genuine kind of change that we know occurs as time passes. This is the change that constitutes the A-series: events from the future coming to be present and then passing away into the past. It is a kind of change we experience and we are aware of.<sup>14</sup> We know from our own experience of ourselves in time that this objective passage exists, and the B-theorist misses this kind of change altogether. The B-theorist may offer back the response that the experience that there is this kind of change is an illusion, and the kind of object-change they offer is the only kind that fundamentally exists. This is a point on which philosophers of time continue to disagree.

## 8.5 The Truthmaker Objection

As we have seen, there are two main packages of views about time: first, the combination of the A-theory and presentism, which has the advantage of making sense of time's passage as we ordinarily experience and understand it, and, second, the combination of the B-theory and eternalism (the block universe view), which seems the most easy to reconcile with special relativity. In this section, we will consider one more reason why some philosophers have worried about presentism and the A-theory.<sup>15</sup> This is the truthmaker objection.<sup>16</sup>

The objection starts from an appeal to truthmaker theory. Truthmaker theorists believe that when a sentence or proposition is true, there must be something in the world that makes it true. The truth of a sentence or proposition cannot be a brute fact. This is the idea that motivates one of truthmaker theory's main theses, **truthmaker maximalism**, the claim that all truths have truthmakers. So, for example, consider the following sentence:

- (1) There are two pandas in the Adelaide Zoo.

**Truthmaker maximalism:** the view that all truths have truthmakers



If this sentence is true, then there must be some thing or things in the world that make it true – these are the sentence's truthmakers. The sentence's truth must be connected with reality in some way, the kinds of objects there are in the world and how they are. Truth cannot “float free” of reality.

Typically, truthmaker theorists think there are particular kinds of entities that may serve as truthmakers. The truthmakers for a sentence consist of the objects the sentence is about having the right sort of categorical features and standing in the right relations to make the sentence true. For example, sentence (1)'s truthmaker will involve two pandas being located within the boundaries of the Adelaide Zoo in Australia at the time the sentence is uttered.

#### TEXT BOX 8.4

##### Categorical vs. Non-categorical Features

When metaphysicians appeal to **categorical features**, they have in mind features that just concern what that object is like at a certain time in those actual circumstances. In other words, an object's categorical features do not concern how the object is (a) relative to other objects, (b) in other possible situations, or (c) at other times. For example, one might describe a particular tennis ball as being spherical or made of rubber. In either case, this would be to ascribe it certain categorical features. Here are some non-categorical features we might ascribe to it:

- being the only yellow object in the room
- possibly being used in the French Open
- having the ability to bounce off of hard surfaces
- having been created in a plant in 2020.

None of these four features concern merely how the object actually is in itself at a given time. The first involves a comparison between the ball and other objects. The next two concern facts about what might be the case in other situations. The fourth involves a fact about the ball at other times. So, examples of non-categorical features are: relational features, modal features, **dispositional features** (features about how an object might behave in various circumstances), and temporal features.

Sometimes we might be interested in the categorical features of collections of objects. For example, we might talk about the categorical features of three tennis balls in a can. These categorical features may include again their being spherical and being made of rubber. But when we talk about the categorical features of collections, these may also include facts about the constituent objects' relations to each other. For example, it is a categorical feature of the three balls in a can that one of these is in the middle of the other two, and also that they are all of the same size. Although these features are relational, they are categorical features of

*the collection* in that they only concern how the collection's constituents taken together are actually and at a given time.<sup>17</sup> They do not concern that collections' relations to other objects, other possible situations, or other times.

Or consider:

(2) Paris is west of Vienna.

This sentence is true. Here, the truthmaker consists of two cities, Paris and Vienna, related in such a way that Paris is west of Vienna.

Similarly, the following two sentences are false because they lack the required truthmakers:

(3) There are two pandas in the San Diego Zoo.

(4) Paris is east of Vienna.

Although there are more than two pandas in the world, there are not two pandas located within the boundaries of the San Diego Zoo. So, sentence (3) lacks a truthmaker. There exist no objects structured in the way the sentence requires for it to be true. Similarly, for (4) to be true, both Paris and Vienna would have to exist and they would have to be ordered in such a way that Paris is east of Vienna. Since Paris fails to be related to Vienna in this way, there is no truthmaker for this truth.

### EXERCISE 8.5

#### Truthmaker Theory

Provide truthmakers for the following truths.

- A. The Eiffel Tower is 324 meters tall.
- B. The average Brazilian female is 1.61 meters tall.
- C. San Francisco is more populous than Seattle.
- D. Glass is fragile.

We have just illustrated truthmaker maximalism using examples of sentences that are about presently existing objects. To see the trouble for the presentist, we need to consider examples of past- and future-tensed truths. Consider the following sentences:

(5) Dinosaurs once roamed the Earth.

(6) The 2028 Olympics will be held in Los Angeles.

Both (5) and (6) are true. So, according to our truthmaker principle, they must have truthmakers. Intuitively, since (5) is about dinosaurs, the truthmaker for (5) would involve the existence of dinosaurs (among other things). The truthmaker for (6) would appear to involve the occurrence of the 2028 Olympics. If one is an eternalist and a B-theorist, then there is no problem accounting for the truth of (5) and (6) since these truthmakers exist. For the eternalist, past and future objects and events are just as real as presently existing objects and events.

Let's be a bit more careful for a moment about how the B-theorist will understand (5) and (6). These are tensed sentences and, recall, the A- and B-theorists have different ways of making sense of tensed statements. As we saw in the previous section, the B-theorist thinks that tensed facts are grounded in tenseless facts. Let's again pretend these sentences are being asserted on May 9, 2023. According to the B-theorist, the fact expressed by (5) is grounded in the fact expressed by  $(5_B)$ :

$(5_B)$  Dinosaurs roam the Earth at some time earlier than May 9, 2023.

And the fact expressed by (6) is grounded in the fact expressed by  $(6_B)$ :

$(6_B)$  The 2028 Olympics are held in Los Angeles at some time later than May 9, 2023.

$(5_B)$  and  $(6_B)$  both express tenseless, B-theoretic facts. For  $(5_B)$  to be true, it must be the case that there exists some time earlier than May 9, 2023, and that dinosaurs are roaming the Earth at that time. For  $(6_B)$  to be true, it must be the case that there exists some time after May 9, 2023, that this be in the year 2028, and that the Olympics are held in Los Angeles at this time. Indeed, we can restate  $(5_B)$  and  $(6_B)$  using the language of predicate logic to say:

$(5_B) \exists t (t \text{ is earlier than May 9, 2023} \wedge \text{Dinosaurs roam the Earth at } t)$   
 $(6_B) \exists t (t \text{ is later than May 9, 2023} \wedge (t \text{ is in 2028} \wedge \text{the Olympics are held in Los Angeles at } t))$

Thus, we can see that as the B-theorist understands (5) and (6), they require the existence of past and future times to be true. If no future or past times exist, then (5) and (6) cannot be true. But they clearly are true. So, given the B-theory, presentism must be false. This argument generalizes. Pick any truths about the past or the future. Given the B-theory of time, the presentist cannot account for their truth.

But, as we have already seen, presentists aren't B-theorists. They are A-theorists. And so, they will not agree that the facts expressed by (5) and (6) are grounded in the facts expressed by  $(5_B)$  and  $(6_B)$ . But then how do A-theorists understand past- and future-tensed truths?

As we have seen, the A-theorist denies that tensed facts are grounded in tenseless B-facts. And this means that the A-theorist would not agree

to use existential quantification over past and future times to represent the logical structure of tensed sentences like (5) and (6). Rather, typically the A-theorist will introduce novel logical notation to capture the phenomenon of tense in symbolic logic.

This tense logic was first developed in the 1950s by the logician Arthur Prior (1914-1969). In tense logic, new operators are introduced that apply to entire sentences.<sup>18</sup> For example, **P** may be used to symbolize the past tense operator. We will read '**PA**' as: In the past, A. **F** may be used to represent the future tense operator. We will read as '**FA**' as: In the future, A. Then we may represent (5) and (6) as:

(5<sub>A</sub>) **P** (Dinosaurs roam the Earth)

and

(6<sub>A</sub>) **F** (The 2028 Olympics take place in Los Angeles)

Now, by itself, the use of tense logic does not assume that the A-theory is correct. What is distinctive about the A-theorist rather is that she believes that the facts expressed by sentences like (5<sub>A</sub>) and (6<sub>A</sub>) cannot be grounded in facts involving quantification over past and future times. The B-theorist has no objection to symbolizing (5) and (6) in the way the A-theorist suggests, using tense operators. The B-theorist will only say that the facts expressed by (5<sub>A</sub>) and (6<sub>A</sub>) are grounded in quantificational facts over past and future times.

We can now return to our main point concerning truthmakers. The presentist as an A-theorist will take (5<sub>A</sub>) and (6<sub>A</sub>) to be the correct ways of understanding (5) and (6), and will claim that these are not grounded in the facts expressed by (5<sub>B</sub>) and (6<sub>B</sub>). But then one wants to know what are the facts that make (5<sub>A</sub>) and (6<sub>A</sub>) true? And what are these sentences' truthmakers? The eternalist who accepts the B-theory can say that it is the existence of other times in which there are dinosaurs and Los Angeles Olympic Games that contribute to making these sentences true, but the presentist must deny this. So, what *are* the truthmakers for (5<sub>A</sub>) and (6<sub>A</sub>) according to the presentist?

Some metaphysicians think there is no good answer for the presentist here and take this to be a decisive reason to reject presentism. We may summarize the objection in the form of an argument:

*The Truthmaker Objection to Presentism*

1. All truths have truthmakers.
2. If all truths have truthmakers and any sentences about the past and future are true, then their truth will require the existence of past and future objects or events.
3. Some sentences about the past and future are true.

4. So, there must exist some past and future objects or events.
5. If presentism is true, no past or future objects or events exist.

Therefore,

6. Presentism is not true.

There are several strategies the presentist may take in responding to the truthmaker objection.

We will consider three distinct strategies for rejecting this argument:

- Reject Premise (1), truthmaker maximalism.
- Reject Premise (2) and claim that there are truthmakers for past- and future-tensed truths that do not involve past and future objects and events.
- Reject Premise (3) and claim that all statements about the past and future are false.

First, is there any reason the presentist can give to reject truthmaker maximalism? We have seen it is motivated by the simple idea that for a sentence to be true, there must be some objects in the world structured in the right way (having the right categorical properties and the right relations) to make it true. However, the principle is not entirely uncontroversial. A common complaint is that truthmaker maximalism is too strong. Consider for example:

Unicorns don't exist.

This sentence is true (sorry), but what is its truthmaker? Non-existent unicorns? Unless we are going to pursue the view of Quine's McX or Wyman considered in Chapter 1, we must insist that there are no such things. Indeed, the sentence seems to be true precisely because there are no unicorns. It is true because something *doesn't* exist. Its truth doesn't require that something *does* exist, as truthmaker maximalism would imply.

And so, the presentist may claim that not all truths have truthmakers.<sup>19</sup> Some sentences can be true without having truthmakers. This opens up some options for what the presentist might say about statements appearing to be about past and future entities (e.g., (5) and (6)). One option is to say that these are brute truths. (5) and (6) are true, but their truth cannot be explained by more fundamental facts.<sup>20</sup> Another option is to say that (5) and (6) are not brute, but that, nonetheless, their truth doesn't require a truthmaker.

One interesting maneuver available to the presentist along these lines would be to appeal to an analogy between past- and future-tensed truths and truths about fiction. For example, consider the sentence:

Harry Potter is a wizard.

This sentence is true, but it isn't true because there is a person this sentence is about and that person is a wizard. What it takes for the sentence to be true is not that there is some person that has a certain property. Rather this sentence is true because of the existence of a certain fiction or story that says there exists a person, Harry Potter, who is a wizard. If so, the logical form of this sentence is not adequately symbolized as:

Wh

Instead, a better way to represent the truth conditions of this sentence is to symbolize it using a new operator. Like the tense operators introduced above, this will be a **sentential operator**, an operator acting on an entire sentence or proposition. We will write ' $F_{HP}$ ' to abbreviate 'In the fiction of Harry Potter ...'

**Sentential operator:**  
an operator acting on  
an entire sentence or  
proposition

$F_{HP}$  (Wh)

The sentence is true just in case it is a fact in the fiction of Harry Potter that Harry Potter is a wizard.<sup>21</sup>

In a similar manner, we may introduce operators relating to truth in other fictions:

'In the fiction of Sherlock Holmes ...'

'In the fiction of *Star Wars* ...'

'In the fiction of *The Hunger Games* ...' and so on.

A presentist may then apply this point about the logical structure of statements about fictions to past- and future-tensed truths. According to the presentist, no past or future times exist. However, she might allow that there is a common fiction with which we are familiar that consists in certain past- and future-tensed claims. In this fiction, it is the case that there are past dinosaurs and future Olympic Games. She could then understand an ordinary assertion:

(5) Dinosaurs once roamed the Earth,

as containing a tacit fiction operator of the form:

(5F) According to the fiction of past entities, dinosaurs once roamed the Earth.

Similarly,

(6) The 2028 Olympics will be held in Los Angeles,

will be thought to express something like:

(6F) According to the fiction of future entities, the 2028 Olympics will be held in Los Angeles.

If this is right, then past- and future-tensed sentences don't need truth-makers - objects they are about - to be true. This would be to presuppose

their truth made some claim about things that genuinely exist. Instead, one might argue, they just require facts about what is true according to a certain fiction. We might call this view temporal **fictionalism**, the view that past- and future-tensed truths are to be understood as truths about fictions.

### EXERCISE 8.6

#### An Objection to Temporal Fictionalism

In this section, we considered the position that tensed truths should be understood as truths about fictions. One might object to this position by noting that important disagreements about the past, about natural history and human history, often turn on whether past-tensed claims are truths or fictions. If one claims all past-tensed truths are really truths about fictions, then one is thereby trivializing these debates. For example, some creationists have claimed that “Dinosaurs roamed the Earth 200 million years ago” is not true, but a fiction. Holocaust deniers claim that “Six million Jews were killed in Nazi Germany” is not true, but a fiction. What is the best response that an A-theorist could give here to defend a fictionalist account of tensed truths?

### TEXT BOX 8.5

#### Fictionalism

**Fictionalism** is a general strategy to which metaphysicians have appealed in order to account for the truth of various sorts of claim, where truthmakers for these claims do not seem available. In Chapter 10, we will consider a position called ‘modal fictionalism’ that takes truths concerning possibility and necessity to involve tacit ‘In the fiction ...’ operators. Fictionalism has also been proposed to make sense of mathematical and moral claims.

For example, in the mathematical case, we noted in Chapter 2 that the sentence ‘There are at least three perfect numbers greater than 17’ appears to quantify over numbers. Adopting the truth of a sentence like this would appear to commit one to Platonism. However, the mathematical fictionalist may argue that the logical form of this sentence is not properly represented as:

$$\exists x \exists y \exists z (((Nx \wedge (Px \wedge Gxs)) \wedge (Ny \wedge (Py \wedge Gys))) \wedge (Nz \wedge (Pz \wedge Gzs))) \wedge ((x \neq y \wedge y \neq z) \wedge x \neq z)),$$

but rather instead as:

$$F_M \exists x \exists y \exists z (((Nx \wedge (Px \wedge Gxs)) \wedge (Ny \wedge (Py \wedge Gys))) \wedge (Nz \wedge (Pz \wedge Gzs))) \wedge ((x \neq y \wedge y \neq z) \wedge x \neq z))$$

where ' $F_M$ ' is an operator that symbolizes 'In the fiction of mathematics ...' The mathematical fictionalist thus thinks of mathematics as a collection of stories similar to the Harry Potter books or *Star Wars*. Just as claims in the fiction of Harry Potter may be true even though Harry Potter does not exist, claims in the fiction of mathematics may be true even though numbers and sets do not exist. What makes these claims true is just that the fiction of mathematics, actual mathematics books and articles, say they are true. (Of course, there may be interesting disagreements about which books and articles count as definitive sources for what is true in the fiction of mathematics. But the same issue arises for what counts as the definitive sources for what is true in the fiction of Harry Potter: just the books or the movies too? Just the books and movies, or the fan fiction too?)

Moral fictionalists deploy a similar strategy to account for the truth of moral claims. Moral fictionalists will deny that there are any features like moral rightness or wrongness that actions may possess. However, they argue that we as humans share in a common fiction that there are such features and that is what allows sentences like, 'That was a morally right action' or 'That was a morally wrong action' to be true.

We have now discussed two ways a presentist who rejects truthmaker maximalism may respond to the truthmaker objection by rejecting the first premise. The first is to claim that the fact that certain past- and future-tensed statements are true is just a brute fact, not capable of any further explanation. The second is to appeal to fictionalism. A third possibility that has been explored is to say that past- and future-tensed truths are true in virtue of the present time having certain basic non-categorical features. For example, a presentist might say (5) is true in virtue of the universe (that is, everything that exists presently) having the property of previously containing dinosaurs. Similarly, she may say that (6) is true in virtue of the universe having the property of in 2028, Olympic Games being held in Los Angeles.<sup>22</sup> Explaining the truth of these claims by appealing to non-categorical features like "previously containing dinosaurs" or "in 2028, having the Olympic Games in Los Angeles" is not a maneuver most truthmaker theorists will accept. As we saw, the truthmaker theorist usually think that all truths may be accounted for in terms of the existence of certain entities possessing *categorical* features and standing in certain relations. These features are not categorical features. As such, the truthmaker theorist will likely think the presentist is cheating if she appeals to the universe's having properties like this to "explain" the truth of (5) and (6). But this is one possible option for the presentist, and one which will allow her to point to things in the world to account for the truth of past- and future-tensed claims.<sup>23</sup>

But let's say that the presentist is sympathetic to truthmaker theory in its standard form. What can she then say to respond to the objection? She



may find another way to respond to the argument by rejecting the move from the argument's first premise to the second. The presentist might argue that (5) and (6) do have truthmakers, it is just that these truthmakers do not consist of past or future objects. All of the truthmakers for all truths, including past- and future-tensed truths, are contained in the present time. Two options are commonly explored.

One strategy is to ground the truth of past-tensed sentences in the existence of traces or records that past objects and events have left behind. Perhaps what makes (5) true is not the existence of dinosaurs, but rather the existence of present dinosaur bones scattered over the Earth. Similarly, the truthmakers for future truths may be present plans and intentions that exist now. What makes (6) true are not future Olympic games, but instead records of meetings of the International Olympic Committee stating that the 2028 Olympics will be held in Los Angeles.

The presentist might worry, however, about having enough traces and records, plans and intentions to serve as truthmakers for all of the past- and future-tensed truths there are. Couldn't there be truths about the past that have left no trace? Perhaps once in 2020 you wore mismatched socks to school, but the socks have since been destroyed, no one photographed this, and the incident has been erased from your memory. Does this mean it is not true that you wore mismatched socks to school that day? The worry is even more striking about future-tensed truths. There are many events in the future of your life, the planet, the universe that will come to pass – many future-tensed truths out there to be known. But so few of these are intended or planned, and even those that are planned may not come to pass. There is thus good reason to look elsewhere for truthmakers for past- and future-tensed truths.

Another option that has been explored is to take the truthmakers for past- and future-tensed truths to be some conditions in the present that necessitate what will happen in the future or what happened in the past. Knowing the present state of the universe and the laws of nature, one might naturally think one could then deduce what all past and future truths are. So then why not take the truthmakers for all past- and future-tensed truths to be the combination of the present state of the universe – all of the objects that now exist and what they are like – plus the laws of nature?<sup>24</sup>

For this to work, there must be some strong constraints on what the laws of nature are like. First, for there to be determinate past and future truths, the laws of nature will have to be deterministic. **Determinism** is the position that the laws are such that given any state of the universe, one can predict with certainty what the state of the universe will be at any other time. There are no chances entering into the laws of nature.<sup>25</sup> In addition, for this response to be compatible with both presentism and truthmaker theory, the facts about the laws of nature must somehow be

**Determinism:** the view that the laws are such that, given any complete state of the universe at a time and facts about what the laws are, it is possible in principle to deduce what the complete state of the universe will be at any later time

grounded in the present. But are there enough truthmakers at the present time to ground what the laws of nature are? One might think that the facts about what the laws are depend on what things are like over extended periods of time.<sup>26</sup>

There is a more fundamental worry, however, about these attempts to respond to the argument by rejecting the move from premise (1) to premise (2). For both ways of carrying this out, whether by appealing to traces and plans or to present states of the universe and the laws, seem to involve a confusion. This is a confusion we have seen before<sup>27</sup>: the mistake of confusing the metaphysical issue of the truth of sentences like (5) and (6) with the epistemological issue of how we could know whether (5) and (6) are true. The discovery of traces and deduction from the laws may very well be ways we have of coming to know about the past and future from what we have to go on in the present. But the issue of what makes sentences like (5) and (6) true is a different issue than the issue of how we could know they are true. To see this, let's go back to the original example used to motivate the truthmaker principle:

(1) There are two pandas in the Adelaide Zoo.

What makes (1) true is the existence of two pandas (Wang Wang and Funi) in the Adelaide Zoo. There is a separate issue of how we might know about this truth. I know about it because I read about the pandas on the zoo's website. Maybe you know it because you have visited the zoo and seen the pandas there and now have a memory trace of seeing them. But these are epistemological issues, not issues about what is (1)'s truthmaker. The truth of (1) is grounded in facts about the pandas themselves, not in memory traces or our ways of knowing about them.

Finally, we can discuss a final strategy the presentist may use to respond to the truthmaker objection. This is to deny the argument's third premise and say that there are no truths about either the past or the future. There are only truths about the present, because all that exists exists in the present. It may be a serious blow to common sense to deny that there are any truths about the past and the future. But the presentist might argue it is sometimes the case that metaphysical inquiry reveals that common sense is wrong. Since past and future objects and events do not exist, neither (5) nor (6) is true.

One immediate worry about denying the truth of all past- and future-tensed truths is what to say about these pairs of sentences.

(5) Dinosaurs once roamed the Earth.

(5\*) Unicorns once roamed the Earth.

(6) The 2028 Olympics will be held in Los Angeles.

(6\*) The 2028 Olympics will be held on the Moon.

Unlike (5) and (6), their starred counterparts seem clearly wrong. But if we deny (5) and (6) are true, aren't we assigning them the same status as (5\*) and (6\*)?

In the end, all that is clear is that the eternalist has an easier time accounting for the truth of past- and future-tensed sentences than the presentist does. This is a mark in favor of eternalism, but certainly not a knockdown argument. Presentists may view this as a challenge to which they will have to work out a response.

### EXERCISE 8.7

#### The Truthmaker Objection

Above, we explored several strategies the presentist might use to respond to the truthmaker objection. Which way seems most promising to you? How would you defend this response against the concerns raised in the text?

## 8.6 Time Travel

We have now seen several distinctions in the philosophy of time: between the A-theory and B-theory of time, between presentism and eternalism. The issue about which of these views is correct is interesting on its own. But, as we will see in this section, these distinctions will also help us understand some other more complicated issues related to time. For example, time travel is a favorite trope in popular books and films, but it is often difficult to think about. There are a host of interesting questions related to time travel. Some of these are scientific or technical questions. Is time travel consistent with the laws of nature? Would it be possible to actually build a time machine? These are certainly interesting questions. But what we will consider here are instead the distinctively metaphysical questions that come up even before we ask about these scientific or technical issues.

To start, one might wonder whether time travel is even logically possible, or whether the very idea of time travel presents one immediately with a contradiction. D.C. Williams once suggested that time travel, if it doesn't just consist in the banal fact that at each moment we occupy a different moment than the one we did previously, must fundamentally involve a contradiction.<sup>28</sup> For example, consider the novel *The Time Machine* in which the protagonist travels in minutes from nineteenth century Victorian England to the year 802,701 AD. We seem asked to imagine that in minutes, he is millennia from where he began. But this appears to be a contradiction. It is impossible to be both minutes away and millennia away

from the same starting point in time. So, time travel, in any interesting sense that would mark the time traveler apart from the rest of us, seems to involve a contradiction.

To resolve the contradiction, philosophers like Paul Horwich and David Lewis<sup>29</sup> have appealed to a distinction between two ways in which we speak of durations of time. Suppose the journey in the time machine takes the time traveler ten minutes according to his experience. Then, as Williams suggests, this time travel will seem to imply a contradiction. The time traveler can say as he steps into the time machine, "Ten minutes from now, I will be millennia from now." To resolve this contradiction, let's first call the sense in which millennia pass **external time**. This is what Lewis characterizes as "time itself." It is how much time has passed objectively according to the world outside the time traveler. On the other hand, we may refer to the sense of elapsed time in which only ten minutes have passed (according to the experience of the time traveler) as **personal time**. This is the time that is measured by the time traveler themselves. It is measured by the ticks of one's wristwatch, the growth of hair on one's head, and the occurrence of other bodily processes. With this distinction in hand, we can resolve the contradiction. When the time traveler steps out of his time machine, ten minutes of his personal time will have elapsed. But millennia of external time will have elapsed. There is no contradiction if we are careful. In general, Lewis suggests we understand time travel as any discrepancy between personal time and external time. In this sense, it seems time travel is at least logically possible.

**External time:** time itself

**Personal time:** elapsed time as measured by the normal behavior of physical objects

There may be no contradiction in the very idea of time travel, but can the time traveler do anything once they arrive in the past? Is it possible to change the past? This has been a vexing issue, but one that may be resolved by deploying the resources of the previous sections. First, let's start by assuming the combination of B-theory and eternalism: the block universe view. If we are asking about travel to the past and assuming there are facts about the past, then the most natural picture to assume in the background is a view in which the past is real and so provides one some locations to travel to.

If the past exists and there are facts about it, then is it possible to change these facts? Well, in a strict sense of "change the past," where we are assuming it was once the case at a particular time  $t$  that an event occurred, and then one brought about a change so that at  $t$ , the event did not occur, the answer would seem to be: no, you can't change the past. Whatever happened, happened. Suppose, for example, that on May 2, 2020, you slipped and fell in the rain. Then, unless you accept the view considered at the end of the last section which denies there are any truths about the past and future, you can't go back to the past and make things so this did not happen.

After all, what if you did succeed in going back to that date and warned yourself so that you do not slip? This would imply a contradiction. This would involve it being the case both that: (a) on May 2, 2020, you slipped and (b) on May 2, 2020, you did not slip. This is a contradiction.

One might concede that whatever one does, it is impossible to undo a fact that occurred and make it so that it did not occur. But one might think there is still a way time could be such that you could go back and stop the slipping on May 2, 2020. This would be if time were not one-dimensional, as we ordinarily think, but instead two-dimensional.

Think of time as represented by a two-dimensional graph with the x-coordinates corresponding to locations in what we ordinarily think of as one-dimensional external time (1950, 2000, 2020, and so on) and the y-coordinates corresponding to alternative timelines, beginning with the original timeline before anyone time travels (see Figure 8.5).<sup>30</sup>

If we look at time two-dimensionally, then we can understand how you might go back to the past and prevent your slipping. When you travel back, you don't change the fact that there was a slipping on May 2, 2020. This slipping still occurs at location  $x = \text{May 2, 2020}$ , and  $y = 1$ . However, what you can do is make it the case that at the location  $x = \text{May 2, 2020}$ , and  $y = 2$ , there is no slipping. That there is a slipping at one location in two-dimensional time and no slipping at a distinct location in two-dimensional time presents no contradiction. But we can now see that even on this two-dimensional model, still, one is incapable of changing the past. The fact remains that you did indeed slip on May 2, 2020, on the original timeline.

It is possible that some of us when thinking about time travel, say in the context of watching a movie like *Back to the Future* where a character does successfully go to the past and change things, may think of time

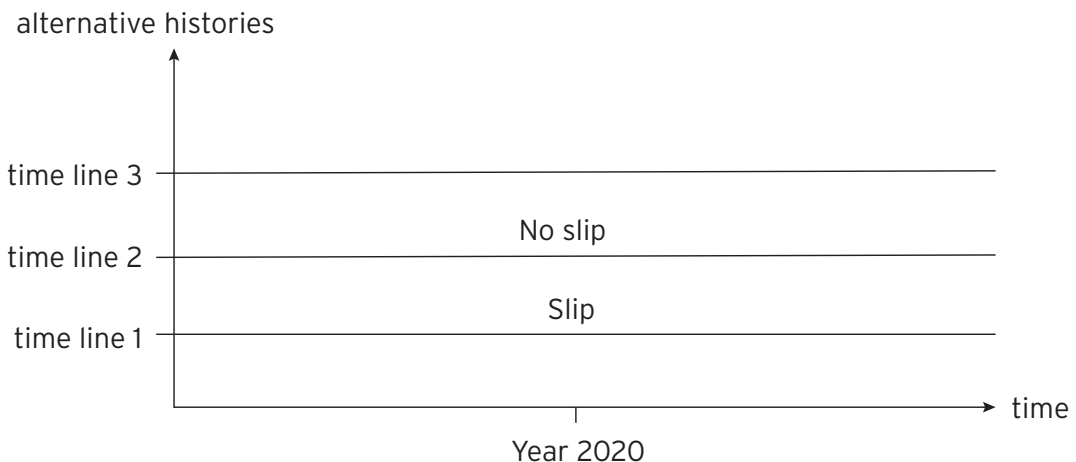


Figure 8.5 Two-Dimensional Time

two-dimensionally. We may be imagining there are multiple timelines. But ultimately this isn't a coherent way to imagine changing the past. For if time is two-dimensional, one doesn't ever go back and change the past. If time is two-dimensional in this way, then one doesn't really ever even go back to one's own past. For on this model, when you go back to prevent your slipping, you are traveling to a different location in time than you were. You don't travel back to (May 2, 2020, 1). You travel to (May 2, 2020, 2). This is a different temporal location altogether. So, when thinking about time and time travel, most philosophers represent time one dimensionally. The two-dimensional model does not seem to help make our thoughts about changing the past more coherent.<sup>31</sup>

No one can undo what has already happened. As we have seen, this would result in a contradiction. But does this mean that if one were able to time travel, one would be frozen, incapable of any action? That one couldn't even step out of the time machine because in doing so one might trample a blade of grass thus generating a contradiction?

No, there is no contradiction in the idea that time travelers may do things in the past, even do an interesting variety of things. As Horwich puts it, even if one cannot change the past, this does not mean one cannot *affect* the past. You can go back to the past and do things. Imagine Smith who in her fascination with Jane Austen decides she will build a time machine, travel to the past and meet her. She can do this. But if this occurs, it will have to have always been the case that she met Jane Austen. There can be no *change* in what happened, no change in the events that lie along the B-series. Nevertheless, the time traveler can certainly be a part of what happened in the past.

### EXERCISE 8.8

#### Time Travel and the A-Theory

Would appeal to an irreducible A-series help to remove the contradiction involved in changing the past? Explain why or why not.

What about changes in one's earlier life? What if you did something in your teen years you now regret? What if Smith always regretted the fact that she never learned to speak a foreign language in high school? Instead, she goofed off in all of her language classes and now as a result, she is monolingual. Finally, at the age of 40, she finds a time machine and travels back to the year 2000 to convince her younger self to pay more attention in Spanish class. Can she do this? Well, certainly there is

no contradiction in Smith's traveling back to the time of her youth and *trying* to convince herself to pay more attention in class. She is certain not to succeed in convincing her younger self, since she can't change what happens and what happens is that she never learns to speak a foreign language. But she can go back and have this conversation with her younger self. If she does this, then, as we have already established, it will always be the case that this happened. But what if she doesn't remember ever having had a conversation with an older version of herself?

There is no changing what happens at any point in the B-series. This would involve a contradiction. However, this leaves open the possibility that Smith did have this conversation with her older self and yet she forgot it. One can go back and affect one's past, that is, play a role in the events that occurred in one's past. However, then it will have always been the case that this is what happened.

We may now finally consider one last case, what is perhaps the most interesting and frustrating one to think about, at least for many metaphysicians. Is it possible to go back in time and kill one's younger self?<sup>32</sup> Here again, logic seems to point to the answer "No." If you could go back in time and kill your younger self, then that would mean you did not survive to the moment in personal time at which you supposedly did the killing. So this would mean that you both killed yourself and did not kill yourself, the latter because you were alive to complete the killing.

So, you can't do it. What's the puzzle? Think about it the following way. It seems you are able to make things quite easy for yourself. You can decide to go back to the time when you were most vulnerable, a time when you were asleep in your crib as a gentle baby, a time at which you know no one else was in your house. Your parents had just crossed the street for an hour to talk to the neighbors. You can make things easier for yourself as well by training. In the months leading up to your time travel, you can train rigorously in what it will take to carry out the murder, practicing with different weapons and techniques. You are able to meditate on the task and deepen your resolve, ensuring that you won't back out at the last moment. You can ensure that when the time comes to commit the act, you are as skilled and remorseless as any assassin. Logic permits you to travel to the past, to enter your childhood bedroom, attack your defenseless baby self. But somehow despite all of this training and how close you are able to come, logic will not permit you in the end to achieve your task. As Lewis notes, in one sense of 'can,' the sense of having the skills it takes, of course you can kill your baby self. However, in another sense of 'can,' the sense of logic, the sense in which what you can do does not imply a contradiction, you cannot kill your younger self. Somehow or other, Lewis points out, you will fail. It seems the power of logic alone must prevent you from succeeding no matter how far you can get.

And this is puzzling. No matter how hard you try, there are some things you cannot do even if you build a time machine and travel to the past. Killing your younger self is one of them.

So far, we have presupposed eternalism and the B-theory. This is indeed the framework in which Williams, Horwich, and Lewis write. Time travel is much more difficult to understand if one presupposes some version of the A-theory with its objective and, on most versions, ontologically loaded sense of time's passage.<sup>33</sup> We have already noted that if one is a presentist, time travel is puzzling since there are no locations in time to travel to other than the one that is immediately next. If one endorses the A-theory, one additionally introduces questions like the following: If I travel to the past, but my friends stay behind and continue to persist into the future, which way does time pass? If any version of the A-theory is true, there must be some objective fact about when the present is. So where does the objective flow of time go, with me or my friends? Either way, there are problems. If time goes with me and my personal time, then my friends cease to exist when I time travel. (Or on a moving spotlight model, my friends are plunged into darkness.) If time continues to pass with my friends' personal time, then by time traveling away from them, I cease to exist. It looks like eternalism and the B-theory is required to even begin to make sense of time travel. However, even then, as we have seen, there are surprising consequences about the limits of one's powers.

## Suggestions for Further Reading

A good place to begin further reading in the philosophy of time is the collection *The Philosophy of Time*, edited by Robin Le Poidevin and Murray MacBeath.

For the argument against presentism from special relativity, see Hilary Putnam's "Time and Physical Geometry." Responses to the problem special relativity poses for presentism may be found in Mark Hinchliff's "A Defense of Presentism in a Relativistic Setting," William Lane Craig's *Time and the Metaphysics of Relativity*, and Ned Markosian's "A Defense of Presentism." James Ladyman's "Does Physics Answer Metaphysical Questions" contemplates the possibility that fundamental physical theories beyond special relativity may not lend support for eternalism.

For a classic defense of the B-theory, see D.C. Williams, "The Myth of Passage." Other contemporary defenses of the B-theory and eternalism may be found in J.J.C. Smart, *Philosophy and Scientific Realism* and the second chapter of Theodore Sider's *Four Dimensionalism*.

For the argument that we have special epistemological reason to favor the A-theory, see William Lane Craig, *The Tensed Theory of Time: A Critical*



*Examination*, chapter 5. See the debate between Dean Zimmerman and Theodore Sider for more on this issue in *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne, and Dean Zimmerman. Simon Keller's "Presentism and Truthmaking" is an excellent discussion of the truthmaker objection. For more on fictionalism, see the essays collected in *Fictionalism in Metaphysics*, edited by Mark Kalderon.

For more on time travel, see the essays collected in *Science Fiction and Philosophy*, edited by Susan Schneider. Two recent books about the metaphysical implications of time travel are *Paradoxes of Time Travel* by Ryan Wasserman and *Time Travel: Probability and Impossibility* by Nikk Effingham.

## Notes

- 1 Thus, according to the common view discussed in this section, time's passage is objective in both of the senses discussed in Chapter 5. Time's passage doesn't depend on human beliefs or other mental states. And time's passage is objective in the sense of being intersubjectively accessible.
- 2 Or to put things more carefully following the discussion of Quine (1948) in Chapter 1, it is not the case that there are any such future people or objects.
- 3 Although we will not focus on the other ontologies of time discussed in Text Box 8.1, the moving spotlight view and the growing block theory, it is a simple matter to generalize the argument to challenge each of these views. This is the topic of Text Box 8.3.
- 4 An excellent primer on the theory of special relativity was written by Einstein himself, Einstein (1916/2013). It includes the thought experiment we will describe momentarily.
- 5 The metaphysician Mark Hinchliff, however, has an interesting paper in which he considers alternative definitions of 'present' that would not make what is present relative to a frame of reference. See his "A Defense of Presentism in a Relativistic Setting" (2000). This issue is also discussed by Sider (2001) and Skow (2015). An interesting position called 'fragmentalism' tries to make sense of the idea of relative existence and tense. See Fine (2005) and Hofweber and Lange (2017).
- 6 Note that this is a different sense in which the future is set than what one means when one endorses the metaphysical thesis of determinism. Above, the claim is just that there are facts about what happens in the future. Whether these facts come to follow from what happens in the present as the result of deterministic or indeterministic laws (see Chapter 12 on Free Will) is a separate matter altogether.
- 7 Recall from Chapter 1 that 'exist' and 'is' are interchangeable.
- 8 For example, one might believe in God or numbers, and believe that these entities exist outside of the spacetime manifold. If so, one will think they exist in the tenseless sense of 'exists.'
- 9 How the eternalist puts this claim will depend on their views about relative fundamentality. This was a topic addressed in Chapter 7.
- 10 This is the same McTaggart thought to be the real life counterpart of Quine's character McX (see Chapter 1).
- 11 McTaggart himself was such a nihilist.
- 12 Another option is to reject fundamentality talk altogether and reformulate the issue using the Quinean framework of paraphrase introduced in Chapter 1.
- 13 This will be understood relative to a particular observer's state of motion. For simplicity, we will leave this tacit in what follows.

- 14 A-theorists like William Lane Craig (2000) and Dean Zimmerman (2008, 2011) have argued that the belief in the passage of time has a special epistemological status, in others words, that we know of the passage of time in a way that is especially direct.
- 15 For more objections, see Sider (2001), Chapter 2. And for replies on behalf of the presentist, see Markosian (2004).
- 16 Although this objection has traditionally been framed using the notion of truth-making, it may be straightforwardly reframed in terms of the other relative fundamentality notions introduced in Chapter 7; for example, by questioning whether facts about past and future objects and events could be ungrounded.
- 17 See also David Lewis on internal relations (1986), p. 62.
- 18 Tense logic was built on the framework of the modal logic that will be introduced in Chapter 10.
- 19 Jonathan Tallant (2009) explores this option.
- 20 Simon Keller (2004) considers this option.
- 21 For more on these sentential operators, see Lewis (1978).
- 22 See, for example, Bigelow (1996).
- 23 For the complaint that this constitutes “cheating,” see Sider (2001), Chapter 2. This is discussed more in Tallant (2009).
- 24 Markosian (2013) explores this option.
- 25 We consider the case for (and against) determinism in Chapter 12.
- 26 This is what is referred to as ‘Humeanism about the laws’: the view that the facts about the laws of nature are reducible to facts about regularities in what happens in our universe over time. See Text Box 12.1.
- 27 See the final section in Chapter 2, on mathematical objects.
- 28 Williams (1951).
- 29 See Horwich (1975) and Lewis (1976). I will follow Lewis’s terminology in this chapter.
- 30 Meiland (1974) considers two-dimensional time in detail.
- 31 The position seems to also be in tension with the physicist’s way of representing time as one-dimensional.
- 32 This is closely related to the “grandfather paradox” - the issue of whether it is possible to go back in time and kill one’s grandfather before one’s parents were conceived.
- 33 For a clever way of showing how the presentist A-theorist too may make sense of time travel, see Keller and Nelson (2001).

# 9 Persistence

## Learning Points

- Introduces perdurantism as a solution to the paradoxes of material constitution
- Distinguishes the main metaphysical views about persistence
- Examines the debate over the existence of temporal parts

## 9.1 The Puzzle of Change

In Chapter 3, we discussed some of the classic puzzles of material constitution, including the Ship of Theseus and the Statue and the Clay. These puzzles caused us to rethink our common-sense views about material objects, given some basic facts about identity, including Leibniz's law.

In the case of the Ship of Theseus, we seemed to have to admit that although both  $S_2$  (the ship whose constitution resulted from slow changes over time with the gradual replacement of planks) and  $S_3$  (the ship constructed from the Ship of Theseus's original, rotten planks) both have good claims to be identical to the original Ship of Theseus, they can't both be the Ship of Theseus since there is more than one of them (and identity is numerical identity). We thus seem forced to choose between the claim that neither is the original ship - the ship has ceased to exist - and the claim that one of the two successor ships ( $S_2$  or  $S_3$ ) is privileged. But what objectively could make one criterion of identity (sameness of material makeup or continuity over time) privileged over the other?

In the puzzle of the Statue and the Clay, we had what initially appeared to be just one object, a lump of clay in the shape of a statue. However, by reflecting on the nature of statues and lumps, we seemed forced to admit that the statue could not be identical to the clay of which it is composed. (The statue and lump differ in their temporal and modal properties and

so, by Leibniz's law, must be distinct.) And so, we seemed committed to the existence of two objects wholly located at the same place at the same time: the lump of clay that we called 'Lump' and the statue, Goliath. There is nothing especially exotic about this case. In general, whenever we consider a material object and the material of which it is made, there will be temporal and modal properties to which one may appeal to show that what may look like one object is really two. We therefore seem to be committed to an abundant multiplicity of objects.

This is how we left things in Chapter 3 more or less,<sup>1</sup> but we are now in a position to consider a new candidate solution to these puzzles. This will allow us to see how we need not choose between identity conditions for our ship. It will allow us to see how both  $S_2$  and  $S_3$  may be later versions of the Ship of Theseus without entailing the identity of  $S_2$  and  $S_3$ . We will also finally see a way to avoid the claim that in the case of the Statue and the Clay there are two material objects wholly present in the same place at the same time, while at the same time acknowledging that statues and lumps of clay generally possess different temporal and modal properties. The solution comes in the form of a theory of persistence for material objects. The thought is, if we understand better what it is for an object to persist over time, we can resolve these puzzles of material constitution.

We'll introduce this theory of persistence next and try to evaluate whether it is indeed the panacea it is often claimed to be. But let's pause momentarily to see why persistence through time has long been thought to be such an interesting issue.

Persistence is now thought of as a distinctively metaphysical problem, but this was not always so. Aristotle, in his *Physics*, asked how objects could persist over time. The core of the puzzle is that as objects persist, this involves their becoming different. Sometimes the change is dramatic, as when a tadpole becomes a frog or an ordinary lump of clay gets transformed into a sublime work of art. But even if nothing particularly extraordinary happens to an object, as it persists through time, it at least changes in one respect: becoming older. Philosophers before Aristotle, such as Parmenides (c.515 BC), responded to this issue by denying that there could ever be persistence through change. For Parmenides, change was impossible and so everything is as it is always and permanently. In his view, if something that is a certain way could become something that is not that way (for example if what is young could become not young, if what is wise could become not wise), then what is could become what is not, which is impossible. What is always is and never is not.

Aristotle's response to Parmenides was that his argument against the possibility of change suffered from a conflation between two senses of 'is.' One may use 'is' to predicate a feature of an object, as when we say a person is young or that she is wise. Or one may use 'is' to speak of an entity's

existence, as when we say the person is, she exists. Once we recognize this distinction between two senses of 'is,' we can see that in allowing that (a) a person that is young may become not young, we are not thereby saying that (b) something that is (exists) becomes something that is not (that does not exist). The latter (b) would indeed imply the destruction of an object rather than its persistence through a change, but this is in no way implied by (a), once we recognize these two senses of 'is.'

There is still the question of how anything may survive a change over a period of time. How could a person that is young survive to become different: something that is not young? Actually, Aristotle recognized several ways changes might occur:

by change of shape, as a statue; by addition, as things which grow; by taking away, as the Hermes from a stone; by putting together, as a house; by alteration, as things which "turn" in respect of their material substance. (Aristotle *Physics*, I.7)

For Aristotle, there must be something that survives through a change. Otherwise, the case will be one of the destruction of one object and coming-into-being of another, rather than the persistence of one continuous object through a change. To explain how there could be continuity through a change, Aristotle proposed the view that substances are complex and consist of two parts: matter and form. For example, in the case of Aristotle, the matter is the flesh and bones, the form is man. In the case of the Eiffel Tower, the matter is iron, the form is that of a tower. This was the core of Aristotle's physics, his theory of the makeup of objects in nature that explains how they are able to change. The view is called **hylomorphism**, from the Greek *hyle* for matter and *morphē* for form. Where there is persistence through a change rather than the destruction of an object, we have continuity of matter through a change in form.

Aristotle's theory has of course been succeeded as a physics. When it comes to understanding the various kinds of changes objects undergo, a lot has been learned in the past 2,400 years. Still, hylomorphic theory remains an almost unseen part of the way we think about material objects, and is still an explicit part of some current metaphysical theories.<sup>2</sup> But even if it is part of a correct total theory of persistence, there are problems for which we need additional conceptual machinery to solve. We need more conceptual tools to take into account the fact that ordinary material objects are at least sometimes able to persist through changes in not just their form but their matter as well. The Ship of Theseus presents one such case. In that thought experiment, we imagined the wooden planks of the ship were replaced slowly over time so that eventually none of the original wood remained. To cite another example, scientists tell us that the cells of a human body are constantly in flux. After approximately seven to ten years, a human body recycles all of its cells.

**Hylomorphism:** the Aristotelian view that substances are complex objects made of both matter (*hyle*) and form (*morphē*)

It looks as if objects like human beings are able to survive through changes even when their underlying matter fails to provide any continuity to explain this persistence. In the next section, we will consider some rival metaphysical theories that attempt to explain persistence. As we will see, one central point of disagreement between these views is whether in order for objects to persist over time, there must be something that exists continuously through the change, as Aristotle assumed when he introduced his matter/form distinction.

## 9.2 Some Views about Persistence

Typically, when we think about persistence, we presuppose that there is some one object that is present both before and after the change. For example, consider our piece of clay, Lump. We may assume there is a certain time at which Lump came into existence. Call that time ' $t_0$ .' As time passes, Lump, that very same object, continues to exist. Perhaps at some later time,  $t_1$ , Lump is rolled up into a ball. Then at some time after that,  $t_2$ , it gets squashed and flattened. Finally at time  $t_3$ , our sculptor gets a hold of Lump and shapes it into a statue of the mythical warrior Goliath. And so Lump persists for the rest of its existence. Our assumption here is that it is the same object, Lump, that exists at each time from  $t_0$  through  $t_3$  and onward. This suggests the picture of the persistence of Lump over time as shown in Figure 9.1.

The view captured in this figure is what is called **endurantism**. It is the view that objects like Lump persist by surviving from one instant to the next, where what survival comes to is strict, numerical identity. Lump persists from  $t_0$  through  $t_3$  by existing at each time along the way from  $t_0$  to  $t_3$ , so that the object (Lump) that exists at  $t_0$  is numerically identical to that which exists at  $t_3$ .<sup>3</sup> If 'persists' is a neutral term that may be used to refer to the phenomenon that each of the rival theories in this chapter is trying to explain, we will say that an object *endures* when it persists in the way the endurantist thinks of persistence. We will next consider endurantism's main rival, the view that objects like Lump persist over time not by enduring, but instead by *perduring*.

**Endurantism:** the view that what persistence amounts to is strict numerical identity over time

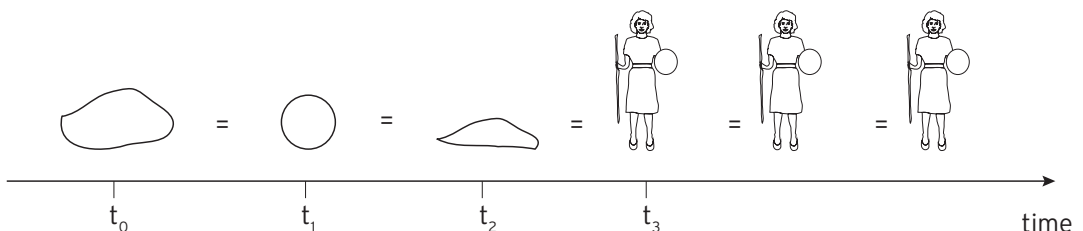


Figure 9.1 The Persistence of Lump (Endurantism)

**Perdurantism (the worm view):** the view that material objects persist by having temporal parts at different times

**Perdurantism** is the view that objects persist over time by being spread out or extended over it. The perdurantist believes that in addition to having spatial parts, material objects also have temporal parts or stages. Material objects are four-dimensional in the sense that they are spread out in time in the same way that they are spread out in space. Just as material objects have different parts at different places (left halves and right halves, arms and legs), they have different parts at different times (first halves and last halves, childhoods and “golden years”). According to the perdurantist, objects like persons or lumps of clay persist over time by having parts (stages) at distinct times. If the perdurantist is correct, then we should not view Lump’s history in the way portrayed by Figure 9.1, but instead view it as depicted in Figure 9.2.

The reason is that, strictly speaking, what exists at  $t_0$  for the perdurantist is not identical to what exists at any later time. For what exists at  $t_0$  is only one temporal part of Lump. What exists at  $t_1$  is a later temporal part of Lump. It is not the same part as that which existed at  $t_0$ . And so on for all other times. The temporal parts of Lump are no more numerically identical to one another than your right arm is identical to your left arm.

Perdurantism is a view that has become increasingly influential since Quine’s presentation of the position in his paper “Identity, Ostension, and Hypostasis” (1950). It is a view, like the eternalism considered in Chapter 8, that is inspired by the idea that we should treat time analogously to space. The eternalist thinks we should treat other times just as we treat other places and view them all as equally real. The perdurantist believes that just as people and other material objects have spatial parts, are spread out over different places, we should view them too as having temporal parts, as being spread out over different times.<sup>4</sup>

To distinguish endurantism from perdurantism, endurantists will often emphasize that their view is not merely that objects persist by existing at

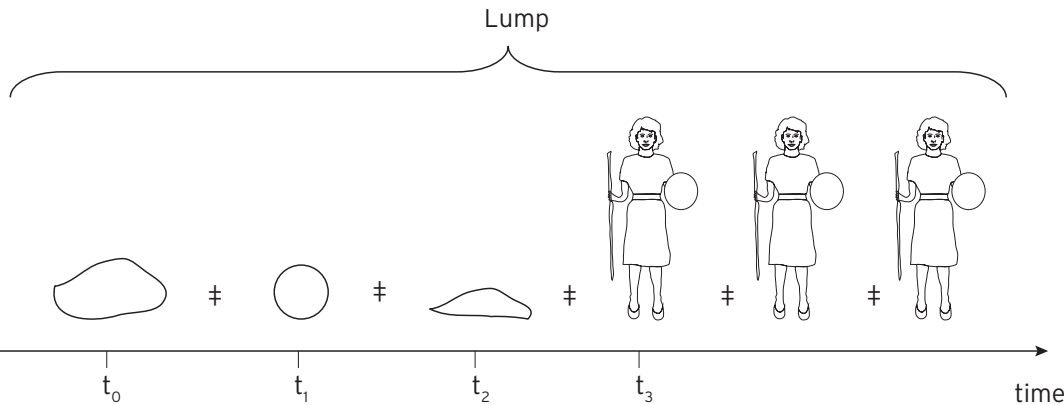


Figure 9.2 The Persistence of Lump (Perdurantism)

distinct times, or even that objects exist by being the same object from one moment to the next. The perdurantist, after all, can agree with both of these claims. The perdurantist can agree that it is the same Lump that exists at  $t_0$  and then again at  $t_1$  and then again at  $t_2$  and so on. For the perdurantist, this is just like if one hid behind a curtain, first poking a hand out, then a leg, then one's head: we would say at each time it was the same person that appeared momentarily from behind the curtain, even though in another sense all that was revealed each time was a part. What the endurantist needs to say to distinguish her view is that objects persist over time by being **wholly present** at each time at which they exist. For the endurantist, it is not just a part of Lump that is there at  $t_0$  and a part of Lump that is there at  $t_1$ . It is all of Lump, the whole object, that is present at each time.

**Wholly present:** to be wholly present at a time is to have all of one's parts exist at that time

Endurantism is typically combined with a related position about material objects: **three dimensionalism**. This is the view that denies any objects have temporal parts in addition to spatial parts. The view is called 'three dimensionalism' because it views material objects as having extension in just the three spatial dimensions: height, width, and depth. Objects also exist in and persist through time of course, but the three dimensionalist does not view this persistence as a way of being extended or spread out over time. Rather, objects are only spread out over space. All of their material parts are spatial parts. By contrast, perdurantists (all of them) will reject three dimensionalism in favor of **four dimensionalism**. This is just the view that at least some objects have temporal parts in addition to spatial parts. Perdurantists are all four dimensionalists because this is required by their view about how ordinary material objects persist over time, by having distinct temporal parts at different times. Later, we will see one way of being a four dimensionalist without being a perdurantist.

**Three dimensionalism:** the view that although objects may have spatial parts, they never have temporal parts

**Four dimensionalism:** the doctrine of temporal parts; the view that in addition to spatial parts, objects have temporal parts

### EXERCISE 9.1

#### Time and Space, Analogies and Disanalogies

We have now seen several metaphysical positions that are based on the view that time and space should be treated analogously. What are some features common to time and space? What are some points of difference? List three of each.

## 9.3 A Solution to the Paradoxes of Material Constitution

Perdurantism is often defended for its ability to help us find solutions to the puzzles of material constitution. So, let's now return to our two puzzles of



material constitution and see how perdurantists have used these puzzles to motivate their position. First, consider the Ship of Theseus. Earlier, we were tacitly assuming endurantism. The question was framed as a question about which object at a later time,  $S_2$  or  $S_3$ , was numerically identical to the original ship,  $S_1$ . This assumed that if the Ship of Theseus was to persist, it must be identical to one of these objects at a later time,  $S_2$  or  $S_3$ . However, we may now see that if we are perdurantists, we won't think that the ship persists over time by there being a later object that is strictly identical to that earlier existing object,  $S_1$ . Instead, we will think that the ship persists by having numerically distinct temporal parts at different times. And so, when we ask about  $S_1$ ,  $S_2$ , and  $S_3$ , the perdurantist will say that none of these objects is identical to any of the others. Rather, these are three, numerically distinct temporal parts. And then the question is, are there any wholes that have both the original object  $S_1$  and either of  $S_2$  or  $S_3$  as parts. And here we must think back to our views in Chapter 3 about parts and wholes.

If we are mereological universalists, as many four dimensionalists are, then we will think that any non-overlapping material objects whatsoever compose some further object. And so, in the case of Figure 9.3, there are at least two temporally extended objects: one that includes  $S_1$  and  $S_2$  as parts, and another that includes  $S_1$  and  $S_3$  as parts.<sup>5</sup> Depending on what is required to be a ship, it is possible to say that both of these temporally extended objects are ships. And if both contain  $S_1$  as parts, then one can say that each is equally worthy of the name 'The Ship of Theseus.' Since there are two ships here, not one, and names are typically understood as terms denoting a single individual (see Text Box 1.4), we might be careful and say that, strictly speaking, we should introduce two names: 'The Ship

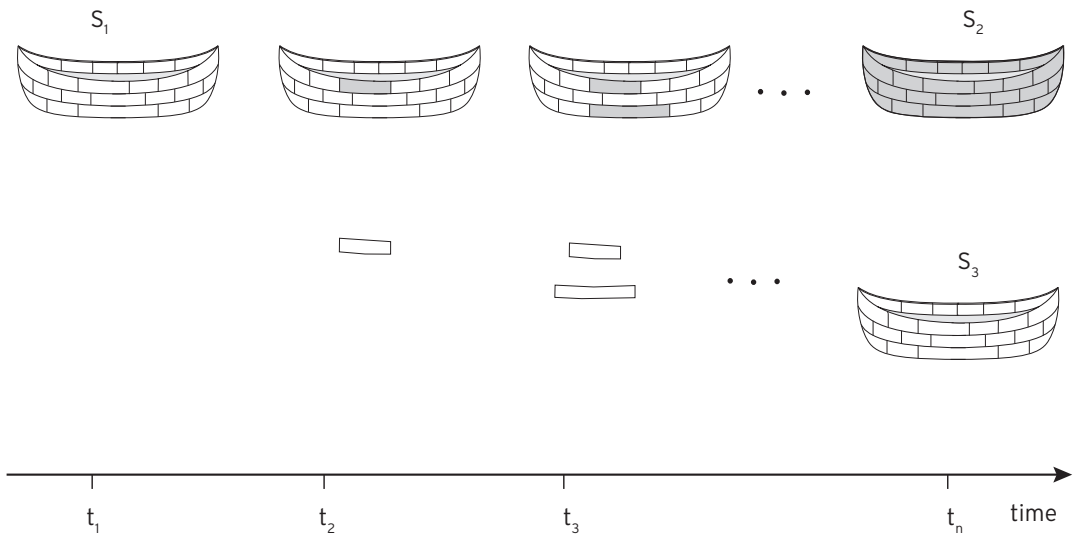


Figure 9.3 The Ship of Theseus

of Theseus-A' and 'The Ship of Theseus-B.' But note that we are now able to say everything we wanted to when we considered the problem.

The problem was that first, we didn't want to have to choose between  $S_2$  and  $S_3$  as the better candidate for being the original Ship of Theseus. Both had good claims to be Theseus's ship. And yet we recognized that if we allowed that both  $S_1 = S_2$  and  $S_1 = S_3$ , then by the symmetry and transitivity of identity, we would be forced to say what is absurd, that  $S_2 = S_3$ . However, if the persistence of the ship does not consist in the strict numerical identity of  $S_1$  with either later object, as the perdurantist thinks, then we don't have to choose between  $S_2$  and  $S_3$  and we are not forced to say the absurd thing.

At this point, one might worry a bit about the perdurantist's solution. For if it is true that both ships contain  $S_1$  as a stage or temporal part, then doesn't this mean that there are really two ships present at the same place at the same time, that there are two ships at the location of  $S_1$  where there certainly would appear to be only one ship?

In a sense, the answer to this question is clearly "yes." If one accepts what has been said up until now, and concedes that  $S_1$  is a part of two ships, then, yes, both ships are present in the same location at our starting time. But this need not be problematic. For first, it isn't true that either ship is *wholly* present at that place at that time. Instead, what is true is that a stage of each ship is present at the same place at that time. Furthermore, this is a stage that each of the two ships have in common. What is happening according to the perdurantist is that the two ships temporally overlap: they share a common temporal part. And the simple reason the perdurantist doesn't find *this* problematic is that it is very similar to other cases we don't find at all problematic: cases of spatial overlap. We have no problem understanding how two objects may share a common spatial part at a time. My body and my left arm share many common spatial parts, including my left hand and my left wrist. When two roads intersect, they share a common part. Sometimes two roads will overlap for an extended stretch. For example, when one is traveling north out of New York City, there are a few miles where I-95 and I-87 overlap. At this time, one may say truly that one is traveling simultaneously on two highways, both I-95 and I-87. But most people (philosophers and non-philosophers alike) find nothing puzzling about this. Once you press the analogy between space and time in the way the perdurantist wants to, temporal overlap (the sharing of temporal parts) seems no more problematic than spatial overlap (the sharing of spatial parts). It is something we can understand and something that obtains in the case of the Ship of Theseus when our two ships share a common stage,  $S_1$ . There are indeed two ships present at  $S_1$  according to the perdurantist. But this need look no more like there are two ships present any more than it need look like there are two roads present at a place where two roads overlap for a stretch.

And now we can turn to what the perdurantist wants to say about the case of the statue and the lump of clay of which it is composed. The perdurantist can note that only an endurantist will be forced to say that this is a case in which two material objects are wholly located at the same place at the same time. For at the time Lump constitutes Goliath, it is true these two objects spatially and temporally overlap, but this doesn't mean these two objects are wholly located in the same place at the same time. For the perdurantist will consider both Lump and Goliath to be objects that are extended over both space and time, and although these objects temporally overlap for some of their histories, they share some stages in common, they do not overlap for their entire histories. Goliath does not come into existence until  $t_3$ . And so they don't share all of the same parts.

Here again, one might try to insist that even though Lump and Goliath are never wholly present at the same spatiotemporal region (because these objects only partially overlap), that these two objects are still both present at the time Lump constitutes Goliath. And then doesn't this imply the Two Object View of David Wiggins we considered in Chapter 3? But as we have seen in the case of the Ship of Theseus, at no time are two material objects wholly present at the same place. For at the time and place of the overlap, there are not two temporal parts, one belonging to Lump and one belonging to Goliath. Rather there is just one temporal part shared by the two objects. Perdurantism allows us to avoid having to say that there are two objects located in that same place at the same time. And we can concede that the statue and lump are distinct (since by Leibniz's law they differ in modal and temporal properties). They are distinct four-dimensional objects.

### EXERCISE 9.2

#### Coinciding Objects

The perdurantist has an elegant solution to the problem of the Statue and the Clay as we have so far presented it. But some philosophers point out that there are variations on this case that pose more of a difficulty for the perdurantist. Suppose the lump of clay and statue are not created at different times, but instead are brought into existence at the same moment. And suppose as well that they are later destroyed at the same moment. What should the perdurantist say about this case? Does this reinstate the appeal of the Two Object View of Wiggins?

## 9.4 The Problem of Temporary Intrinsics

As we have noted, perdurantists take it to be a virtue of their position that they are able to solve what were earlier thought to be problem cases

involving material constitution. However, this is not the only argument for the view. The primary argument for the view, at least so far as some perdurantists are concerned, comes from considering the **problem of temporary intrinsics**. This problem, a problem for endurantism, was first stated by David Lewis in his *On the Plurality of Worlds* from 1986. Lewis calls it the “principal and decisive objection against endurance,” and in favor of perdurantism.

To see the problem, note that among the different properties that material objects possess, some of these appear to be **intrinsic properties**. When we speak of intrinsic properties, we mean properties objects have just in virtue of how they are in themselves, not how they are in relation to other things. Lewis takes shapes and sizes to be paradigm cases of intrinsic properties. When we speak of an object being spherical or a cube or we say it has a certain volume, we are not thereby relating it to other objects that exist, as we might if we said the object was a father or married or that it was located on the surface of the moon. As Lewis notes, many of the intrinsic properties that objects possess are had only temporarily by the objects that instantiate them. To use his own case, it happens at certain times that he (Lewis) is standing and thus has a straight shape; it happens at other times that he is sitting and so has a bent shape.

But this raises a problem for endurantism.<sup>6</sup> Recall that the endurantist thinks that when an object persists from a time  $t_1$  to a later time  $t_2$ , this is because there is one object wholly present at both times  $t_1$  and  $t_2$ . Let's consider the case in which Lewis changes his shape by sitting down. Call the object that is wholly present at  $t_1$ :  $L_1$ . Call the object that is wholly present at  $t_2$ :  $L_2$ . According to the endurantist, if Lewis persists from  $t_1$  to  $t_2$ , if he can persist through this change in shape, then this will be because  $L_1 = L_2 = \text{Lewis}$ . But if  $L_1 = L_2$ , then it cannot be the case that Lewis is straight at  $t_1$  and bent, i.e., not straight, at  $t_2$ . This would entail a violation of Leibniz's law. The same thing cannot both have and lack the same property. So if endurantism is true, then this means that no persisting object can ever undergo a change in its intrinsic properties. This would always involve a violation of Leibniz's law.

Lewis considers a couple of ways an endurantist might try to avoid this problem, although neither of them is, he thinks, satisfactory. One thing an endurantist might do is endorse presentism, the view considered in Chapter 8 that only present objects and events are real. If presentism is true, then we may avoid contradiction by saying that actually nothing is ever both straight and not straight, even if endurantism is true and objects change their intrinsic properties. The presentist will say that at  $t_1$ , Lewis is straight and not bent. Lewis *will be* bent when he sits down. But he is not bent since only what is present exists. And the presentist can say at  $t_2$ , there is also no tension with Leibniz's law, since at this point Lewis is bent and not straight. Lewis *was* straight, but he isn't straight. As it turns out, many endurantists are presentists (even if not primarily for this reason!)

**Problem of temporary intrinsics:** a problem for endurantism that it cannot account for change in an object's intrinsic properties

**Intrinsic properties:** properties objects have just in virtue of how they are in themselves, not how they are in relation to other things

and so this is indeed a response to the problem that will be attractive to many an endurantist.

This is not something Lewis himself finds attractive. He says this is a solution that:

rejects persistence altogether . . . in saying that there are no other times . . . it goes against what we all believe. No man, unless it be at the moment of his execution, believes that he has no future; still less does anyone believe that he has no past. (Lewis 1986, p. 204)

The presentist will of course reject the implication that the view goes against what we all believe. She may concede that it is true that there exists no future or past moments. But she does of course allow that many of us *will* live to do many things. (In the future, a different moment will be present.) And she will allow that for all of us that we *did* exist. (Again, these times are now unreal, but were real when we were living them.) And this should be enough to capture what we ordinarily mean when we say we have a future or a past.

A second endurantist response Lewis considers involves maintaining eternalism, but instead reinterpreting how we think of properties like shapes. This endurantist will say that we might have thought that shapes were intrinsic properties, but really these properties hide more fundamental relations. Objects only have shapes relative to times. And so, strictly speaking, Lewis does not ever have the intrinsic property of being straight; there is no such property. Instead, Lewis has the relational property of being straight-at- $t_1$ . He also has the relational property of being bent-at- $t_2$ . He may therefore lack the property of being straight-at- $t_2$ , but this doesn't entail that he lacks the property of being straight-at- $t_1$ , and so there is no conflict with Leibniz's law.

There is much to say about this response and, again, this is a response that has been taken up by endurantists to solve the problem of temporary intrinsics. Lewis himself calls the view "incredible," saying, "if we know what shape is, we know that it is a property not a relation" (1986, p. 204).

There is certainly something to be said on Lewis's behalf. It seems like an object's shape is just a fact about it. It doesn't seem at all to involve a relation to a time, like the property of being late or beginning at noon. But maybe we can concede that shapes are intrinsic properties while at the same time allowing that they may involve relations to times. This is a possibility Sally Haslanger has explored. To see Haslanger's proposal, start by understanding a "monadic property" as one that is best represented in logical notation as ' $Mx$ ,' as opposed to as ' $Mxy$ ,' as we typically understand relations. It is a property that an object has on its own, but not in virtue of standing in a relation to another object. Haslanger argues:

Relations to times are exactly the sort of relations that may plausibly count as intrinsic. For example, consider two balls,  $b$  and  $b^*$ , that are

intrinsic duplicates. Plausibly intrinsic duplicates can exist at different times, so suppose *b* exists at *t* and *b*\* exists at *t*\*. Now suppose *b* and *b*\* differ in their relational properties, e.g., *b* is red at *t*, but *b*\* is not red at *t*\*, or *b* is 3 inches in diameter at *t* but *b*\* is not 3 inches in diameter at *t*\*. Surely, contrary to our original supposition, we should not count the balls as intrinsic duplicates even if they only vary in the relational ways just indicated; but if the balls must be alike in certain relational respects in order to be intrinsic duplicates, then it is plausible to say that their intrinsic nature is not captured by their monadic properties. Conversely, suppose that no temporary properties are monadic (namely, all temporary properties are relations to times), but *x* and *y* stand in all the same two-placed relations to their respective times (so where one is red at *t*, the other is red at *t*'; where one is 3 inches diameter at *t*, the other is 3 inches diameter at *t*', etc.). Is it not plausible that they are intrinsic duplicates? (Haslanger 2003, p. 330)

So, there is a case to be made, even if one concedes that the endurantist must reinterpret shapes as relations to times, that shapes and sizes and some other temporary properties are still, in an important sense, intrinsic properties.

As we have noted, Lewis prefers neither of these responses to the problem of temporary intrinsics. A much more straightforward response in his view is to give up on endurantism altogether and embrace perdurantism. If we are perdurantists, then there is no problem in saying that Lewis is straight at *t*<sub>1</sub> and not straight at *t*<sub>2</sub>, because what this really comes to for the perdurantist is Lewis having one temporal part that is straight and a distinct temporal part that is not straight. There is thus no violation of Leibniz's law. Shapes are intrinsic properties (and not relations to times), but they are intrinsic properties of temporal parts. Objects like us can persist through change in our intrinsic properties because, according to Lewis, only part of a person exists at any given time and there is no incoherence in an object having one part that has a feature and another part that lacks that feature. My right arm may be bent, while my left arm is straight.

This is a solution to the problem of temporary intrinsics that many are happy with. It doesn't involve relativizing properties to times or denying eternalism. But some four dimensionalists have pointed out that there may be a better way to solve the problem. This is by embracing an alternative four dimensionalist theory of persistence: what is known as the stage view or **exdurantism**.

The exdurantist, like the perdurantist, believes in the existence of temporal parts. He believes in all of the material objects the perdurantist does. However, he will note, if one wants to capture the idea that *it is Lewis* who is bent at one time and straight at another, the perdurantist does not

**Exdurantism (the stage theory):** the view that familiar, material objects are identical to temporary stages

succeed. Lewis after all is not really ever bent or straight on this view. It is only his temporal parts that have these features.

The exdurantist argues that Lewis is not to be identified with the temporally extended material object that *has* parts that are bent or straight. Instead, according to the exdurantist, Lewis *is* one of these parts that is bent or straight. Lewis is a temporary stage. Which stage the name 'Lewis' refers to will depend on features of the context in which the name is being used. Theodore Sider originally defended this position in a paper titled "All the World's a Stage."

Note that the difference between the perdurantist and exdurantist is not ontological. In principle, both perdurantists and exdurantists can hold the same view about composition. If so, they will believe in all of the same entities and agree about which entities are temporal or spatial parts of which others. The difference between the two views has only to do with which of these entities are the ordinary, familiar objects we ordinarily think about and have names for, like 'David Lewis' or 'the Eiffel Tower' or 'the planet Venus.' The perdurantist thinks that objects like this are temporally extended objects possessing temporal parts. The exdurantist thinks they are temporary stages.

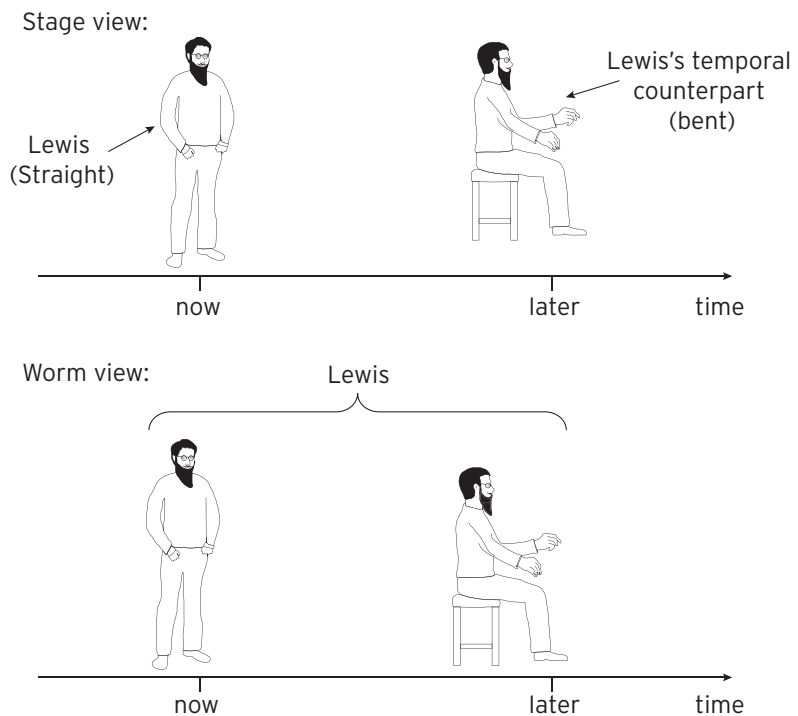


Figure 9.4 The Stage View and the Worm View

**TEXT BOX 9.1****The Stage View vs. the Worm View**

Exdurantism is called the stage view because it identifies the familiar material objects we ordinarily think of as persisting (what are sometimes called 'continuants') with temporary stages. People, statues, dogs and cats, stars and planets: these are all stages, according to the exdurantist. To distinguish it from the stage view, perdurantism is often called the worm view. This is because the perdurantist thinks of ordinary material objects like people or statues as four-dimensional spacetime worms extended not just over space, but over time as well.

**EXERCISE 9.3****Exdurantism and the Problem of Temporary Intrinsic**

Do you agree with Sider that the exdurantist provides a better solution to the problem of temporary intrinsic than the perdurantist? Is Lewis forced to admit that when one says "Lewis is straight at  $t_1$ " or "Lewis is bent at  $t_2$ " that what is said is literally false? Explain your answer.

**9.5 Exdurantism**

That it is supposed to do a better job at solving the problem of temporary intrinsic is one reason Sider thinks we should prefer exdurantism to perdurantism or endurantism. Sider also thinks the view does better at explaining what to say about cases in which spacetime worms overlap. Recall in the case of the Ship of Theseus, the perdurantist was forced to say that in a real sense there are two ships present at the initial time.  $S_1$  is a part of two ships, the Ship of Theseus-A (that contains  $S_2$  as a later part) and the Ship of Theseus-B (that contains  $S_3$  as a later part). And so, if we ask at that time, how many ships are present, according to the perdurantist, it would seem the right answer is "two." However, Sider notes, both sides concede (indeed it is an important part of the perdurantist's view) that there is only one stage there,  $S_1$ . If ships are not extended spacetime worms, but instead temporary stages, as the exdurantist argues, then the



answer to “How many ships are present?” at that time will be “one.” And this seems to be a better answer.

So, there are some reasons to think exdurantism might do a better job even than perdurantism at solving the puzzles of change we have discussed in this chapter. One might nonetheless have qualms about exdurantism. Primarily, there is a question about whether it even constitutes a genuine theory of persistence. Persistence seems by definition to involve an object’s existing at multiple times. The endurantist and perdurantist offer rival views about what this comes to. But the stage theorist, in identifying familiar objects with temporary stages, would seem to deny that any of the objects we are ordinarily interested in discussing ever persist over time.

In response to this concern, the exdurantist typically adopts the strategy of semantic ascent (see Text Box 1.3). Strictly speaking, her metaphysical view does entail that ordinary objects like people, buildings, and planets exist only momentarily. However, she can give a theory that explains why sentences we ordinarily assert that appear to imply these objects’ continued existence over extended temporal durations are nonetheless true.

Consider the following example. The planet Venus exists now. Most of us believe that the planet Venus has existed for many millions of years and will continue to exist in the future. But if one is an exdurantist, then strictly speaking one has to deny this. Strictly speaking it is false that the planet Venus has existed for millions of years since the planet Venus, according to the exdurantist, is a temporary stage. Nonetheless, the exdurantist is able to offer us a paraphrase of our ordinary claims that appear to assert otherwise. We may say, for example:

(1) The planet Venus will survive for the next hundred years.

Here is how the perdurantist and endurantist will understand what this comes to:

(1<sub>Perdurantist</sub>) The planet Venus has a temporal part that exists a hundred years from now.

(1<sub>Endurantist</sub>) The same entity, the planet Venus, that exists now will exist a hundred years from now.

The exdurantist thinks that both of these claims are strictly speaking false, but what is true is the following:

(1<sub>Exdurantist</sub>) There exists a stage a hundred years from now that bears the temporal counterpart relation to the planet Venus.

And this is a better way to understand what we mean when we assert our original sentence (1).

According to the exdurantist, the planet Venus is a temporary stage. But the exdurantist does think there are other stages in the future and in the past, many of which bear interesting relations to Venus. These other stages are similar in many ways to Venus – have the same size, shape, physical makeup, velocity, position. In addition, these other stages bear causal relations to Venus. The past stages are causally responsible for the way Venus is now. The later stages are affected by what Venus does now. If one were a perdurantist, one would say this is because these earlier and later stages are temporal parts of Venus. The exdurantist says this instead makes these earlier and later stages temporal **counterparts** of Venus. They are not parts of Venus itself, but instead distinct objects that bear salient similarity and causal relations to Venus. And it is in virtue of the existence and nature of these counterparts that sentences like (1) are true.

**Counterpart:** a counterpart of one entity x is an entity that bears some salient similarity and causal relations to x

What goes for the planet Venus goes for other familiar objects as well. According to exdurantism, you too are a stage. And this means that strictly speaking, your childhood memories concern things that happened to other people, other stages in the past. Still because these other stages bear salient similarity and causal relations to you, they are important to you in various ways. You are the way you are today because of the way they were then; they are your counterparts. And so, if one of these stages took piano lessons, it is in virtue of this fact that the following sentence will be true:

(2) You took piano lessons when you were young.

And if a future counterpart of yours will take a trip to Israel in the next twenty years, then it is in virtue of this fact that the following sentence is true:

(3) You will travel to Israel at some time in the next twenty years.

According to the exdurantist, people are stages, and as stages they only exist for an instant. Still the exdurantist has a particular metaphysical explanation available for sentences we assert that look superficially to require a person's extended existence over a period of time. So, although strictly speaking for the exdurantist, people and other familiar objects are temporary, he will say it is true that they persist. Facts about their persistence are grounded in the existence of counterparts in the future and the past.

#### EXERCISE 9.4

##### Exdurantism and Temporal Counterparts

According to the exdurantist, could it be true that any person is 30 years old? How could this be if the exdurantist believes we are all temporary stages?

## 9.6 Defending Three-Dimensionalism

For most of this chapter, we have considered the case in favor of the two four dimensionalist views, perdurantism and exdurantism. But is there anything to be said for endurantism in its defense? Endurantism is frequently characterized as the commonsense view on the nature of persistence. Before we think too hard about the puzzles of material constitution, about Leibniz's law, and temporary intrinsics, this view may certainly seem plausible: that the me that exists now, just at this time right here, is the same as, is strictly and numerically identical to, the me that existed yesterday and the day before. Many philosophers will concede that common sense counts for something in metaphysics. What we will examine here is whether there is anything more to be said in favor of endurantism.

Indeed there is. What we will do here is outline two interesting ways in which metaphysicians have defended endurantism against its four dimensionalist rivals. The first way involves calling into question the very notion of a temporal part of a material object. Perdurantists believe that objects persist over time by possessing temporal parts at each time they exist. Exdurantists (stage theorists) identify ordinary, familiar objects like people and planets with temporary parts of larger objects. If one denies material objects ever have temporal parts, one is thereby denying four dimensionalism in either form.

Peter van Inwagen and David Wiggins are two metaphysicians who have denied that it makes sense to think of material objects as possessing temporal parts. To one seeing this debate for the first time, one might be confused how this could be. Isn't the analogy with spatial parts sufficient to make this clear? What could these philosophers not be understanding? Let's consider Wiggins first.

Wiggins, in his *Sameness and Substance Renewed* (2001), insists that four dimensionalism is confused. For Wiggins, it is not even an answer to the question we are interested in about how objects persist over time. The conceptual problem for Wiggins isn't in understanding the notion of a temporal part per se. Rather the problem is in understanding the notion of a temporal part of a material object. Wiggins has no problem conceding that other types of entities, in particular, events, have temporal parts. He argues:

An event takes time, and will admit the question 'How long did it last?' only in the sense 'How long did it take?' An event does not persist in the way in which a continuant does – that is *through* time, gaining and losing new parts. A continuant has spatial parts. To find the whole continuant you have only to explore its boundaries at a time. An event has temporal parts. To find the whole event you must trace it through the historical beginning to the historical end. (Wiggins 2001, p. 31)

This passage is vivid. You may imagine examining any object, your cell-phone, your best friend, and asking what are its boundaries, where are they wholly located. If it is easy to see how you might at a time examine one of these objects and feel satisfied that you have located its boundaries, even though you have only examined it at one time, then you understand Wiggins's point. If all of an object's boundaries can be discerned at one time, then it cannot be an object extended in time, for then it would have more parts, and a boundary that you cannot see.

In his paper "Temporal Parts and Identity across Time" (2000a), van Inwagen expresses puzzlement as well about attributing temporal parts to material objects. Van Inwagen doesn't claim to have a problem understanding how there could be what he calls a 'person stage.' There could be a person who exists just for a short time and then goes out of existence. What he doesn't understand is how such things could be parts of ordinary people like you or me:

God could, I suppose, create *ex nihilo*, and annihilate a year later, a human being whose intrinsic properties at any instant during the year of its existence were identical with the intrinsic properties of, say, Descartes at the "corresponding" instant in, say, the year 1625. And if God could do that, he could certainly create and annihilate a second human being whose one-year career corresponded in the same way to the 1626-part of Descartes's career. But could God, so to speak, lay these two creations end-to-end. . . . Well, he could create, and two years later annihilate, a human being whose two-year career corresponded to the 1625/1626-part of Descartes's career. . . . What I cannot see is how, if God did this, it could be that the "two-year-man" would have first and second "halves." More exactly, I don't see how it could be that the first half of a two-year-man's career could be the career of anything, and I don't see how it could be that the second half of the two-year-man's career could be the career of anything. When I examine the story of the creation and annihilation of the two-year-man, I don't find anything that comes to the end of its existence after one year: the only thing "there" (as I see matters), the two-year-man, will not come to an end after one year; he will rather, continue to exist for another year. And, in the same way, when I examine the story, I don't find anything in it that begins to exist halfway through the story. (van Inwagen 2000a, p. 446)

This story begins to capture what troubles van Inwagen about the existence of temporal parts. There may be temporarily existing objects and then there are stages in a sense, but this doesn't entail any temporal *parts*. When considering a person, whether one who lives for one, two, or fifty years, there is no thing in its spatial location that only exists for part of their lifetime. There is only the person themselves.

**TEXT BOX 9.2****Extended Simples**

Throughout this chapter, we have sometimes spoken of objects *being extended* through space and/or time and objects *having parts* in various places and/or times. It is fine generally, when we are talking about ordinary, material objects, to run these two notions together. However, it is worth briefly mentioning that to say an object is extended (through space or time) is not the same as to say that it has spatial or temporal parts. To say that something is extended in space or time is to say that it occupies a greater than point-sized location in space or time. Tennis balls, which are spheres with diameter 6.7 centimeters, are extended. Electrons, when they are assumed to be point particles, are not extended.

To see that there is a genuine distinction here, we may note the conceptual possibility of objects that are (spatially) extended simples. These are entities extended in space, but lacking spatial parts. Ordinarily, when we think of objects that are spatially extended, we think of them as having parts. For example, tennis balls are made up of bits of rubber and felt. However, it is conceivable that there be spatially extended objects that lack parts. As Kris McDaniel has pointed out in his essay “Extended Simples” (2007), the strings of string theory appear to be examples of extended simples. These are postulated to be fundamental physical objects, not made of any smaller parts. And yet they are claimed to have spatial extension, extension not just in the familiar three dimensions of our ordinary experience, but in additional “hidden” dimensions as well.

**EXERCISE 9.5****Van Inwagen’s Argument Against Four Dimensionalism**

Reconstruct van Inwagen’s argument in numbered premise form. Which part of this argument would you object to if you wished to defend perdurantism?

So, on the one hand there are conceptual worries about the very thought that objects like people and tables have temporal parts. There are also worries about the four dimensionalist’s ability to provide a compelling and coherent account of persistence and change. One thing some three dimensionalists press, something one finds in Wiggins and also the work of Haslanger, is that in denying strict identity over time, the four dimensionalist is denying the whole phenomenon of persistence in the first place. For is an object really *persisting* over time if there is not one, numerically the

same thing that is present at one time and then also present at a later time? To come back to the discussion in the first section, what the four dimensionalist theories present us with appears more like the destruction and generation of a succession of temporary objects, rather than the continued persistence of some one object over time. Since temporal parts are confined to instants, they are not strictly identical from one time to the next. And so they do not persist. But then nothing persists according to the four dimensionalist.

We have seen the exdurantist's response to this problem. They semantically ascend and provide an account of the truth conditions of sentences about persistence in terms of the future or past existence of an object's temporal counterparts. Perdurantists also have a response. They think it is in virtue of objects' possessing temporal parts at successive times that they persist. In addition, the perdurantist will note that it is true for her, just as it is for the endurantist, that there is one thing that survives in the strict sense of being identical to itself over time - this is a temporally extended, four-dimensional spacetime worm. Whether this is enough to satisfy the endurantist is another matter.

## Suggestions for Further Reading

For two defenses of four dimensionalism, see Sider's *Four Dimensionalism* and Mark Heller's *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter*. The collection, *Persistence: Contemporary Readings*, edited by Sally Haslanger and Roxanne Marie Kurtz contains many essential readings. Katherine Hawley's article, "Temporal Parts," in the *Stanford Encyclopedia of Philosophy* is another excellent resource and contains an extensive bibliography. For discussion of an interesting problem about material objects that coincide for all of their histories (not just for part of their histories as in the case of the statue and the lump described in this chapter), see Karen Bennett's influential paper, "Spatio-Temporal Coincidence and the Grounding Problem." For more on intrinsic properties, see "Defining 'Intrinsic'" by Rae Langton and David Lewis.

## Notes

- 1 We did see how denying the existence of most of these material objects, as the mereological nihilist or van Inwagen does, will allow us to escape the problems.
- 2 Most explicitly, this is a key component of the theory proposed in Kathrin Koslicki's *The Structure of Objects*.
- 3 Perhaps Lump does not have to exist at *each* time along the way from  $t_0$  to  $t_3$  in order to persist from  $t_0$  to  $t_3$ . Perhaps objects can have *gappy* histories, flickering in and out along the way. This is a metaphysical issue we will set aside in this chapter.
- 4 To say that something is extended in space is not the same as to say it has parts in different spatial regions. Similarly, we should note that to say some thing is

extended in time is not the same as to say it has parts at different times. Still, since in most cases what is extended in space or time has parts at different spatial or temporal locations, we will move back and forth between talk of extension and talk of mereological complexity. See Text Box 9.2 on extended simples for more discussion. The essays in the 2014 volume *Mereology and Location*, edited by Shieva Kleinschmidt, also address this distinction.

- 5 There will additionally be an object that includes all three of  $S_1$ ,  $S_2$ , and  $S_3$  as parts, but since this is not a candidate for being a ship (since at certain times it is the mereological sum of two ship-stages rather than a single ship-stage), we will ignore this object.
- 6 As we saw in the first section, the fact that objects may come to change their properties is a classic problem recognized by the ancient Greek philosophers. Lewis turns this into a distinctive problem for the endurantist by focusing on the case of objects' changes in their *intrinsic* properties.

# 10 Modality

## Learning Points

- Introduces the modal notions of necessity, possibility, and contingency
- Introduces the distinction between de re and de dicto modality
- Presents several reductive theories of modality
- Evaluates the prospects for essentialism, a thesis about de re modality

## 10.1 Possibility and Necessity: Modes of Truth

In this chapter, we examine modal claims. **Modal claims** express facts not about what merely happens to be the case or what things are actually like, but involve the notions of possibility or necessity. They concern what is possible or impossible; what is necessary or contingent (where something is **contingent** if it is neither necessary nor impossible).

We use the concepts of possibility and necessity all of the time. These concepts are key components of many metaphysical claims as well. But on the face of it, it isn't clear what claims involving these notions mean and in virtue of what they may be true or false. Modal claims don't describe how things actually are, so how can they express facts about our world? Wittgenstein famously claimed that the best way to understand necessity was to see necessarily true claims as nonfactual, as not expressing facts about how the world is, but instead stating our conventions about what we are disposed to take as irrefutable. In this chapter, we will encounter some of the more influential ways of understanding modal claims in contemporary metaphysics. Most metaphysicians want to hold on to the factuality of claims about what is necessary and possible. They then attempt to reduce modal statements to statements that are more easily capable of being understood.

**Modal claim:** a claim that expresses a proposition about what is possible, impossible, necessary, or contingent

**Contingent:** neither necessary nor impossible



**TEXT BOX 10.1****Modes of Truth**

Modal claims concern modes of truth or falsity. A proposition or sentence may be:

*possibly true*, for example, that Abigail Adams was the first president of the United States,

*impossible*, or *necessarily false*, for example, that  $2+2 = 5$ ,

*contingent*, or *both possibly true and possibly false*, for example, that George Washington was the first president of the United States,

*necessarily true*, for example, that  $2+2 = 4$ .

**10.2 Varieties of Possibility and Necessity**

There are several senses of possibility and necessity in which one might be interested. For example, we might ask:

Is it possible to build a vehicle that travels faster than the speed of light?

Philosophers generally recognize at least two senses of the word 'possible' that would give two different answers to this question. One sense is: possible according to the laws of nature, or **nomologically possible** (from the Greek 'nomos' for law). If the laws of nature don't rule out that a certain proposition  $p$  is the case, then  $p$  is nomologically possible. In this sense of 'possible,' the answer to our question is "no." It is not possible to build a vehicle that travels faster than the speed of light. This is something that the laws of nature, special relativity in particular, rule out.

On the other hand, there is another sense of 'possible.' Here, a proposition is possible just in case it doesn't itself entail any contradiction. This is what is often called **logical possibility**. This was the sense of 'possible' we used when we defined the notion of deductive validity and said that an argument is valid just in case it is not possible for its premises to all be true while its conclusion is false. It is logically impossible that  $2+2 = 5$  or that there are triangles that have four sides. It is logically impossible as well for there to be any round squares, married bachelors, or dogs that are not dogs.' In this sense of 'possible,' cars traveling faster than the speed of light are possible. Even if such technology is incompatible with the laws of nature, and so physics would have to be different for there to be cars traveling faster than light speed at our world, there

**Nomological possibility or necessity:** possibility or necessity according to the laws of nature

**Logical possibility:** what does not entail any contradiction

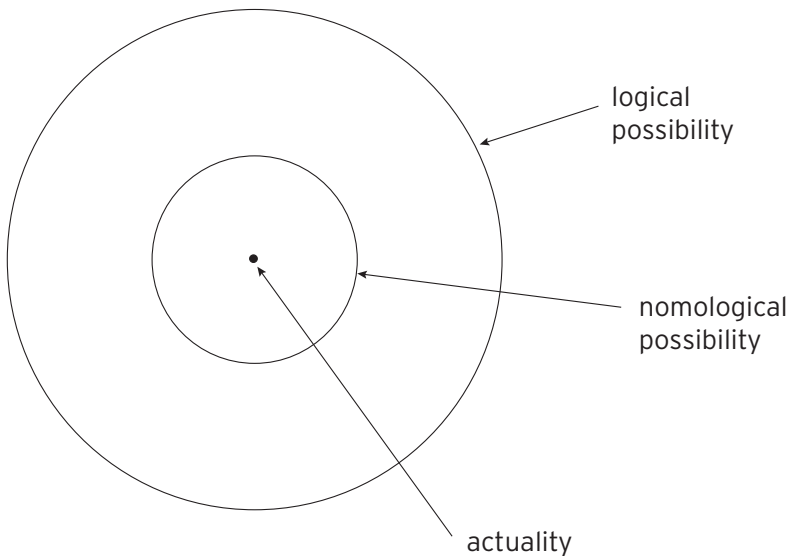


Figure 10.1 Nomological and Logical Possibility

is no contradiction contained in the very idea of a car that can travel at superluminal speeds. This is ruled out by the laws of physics, but not by the laws of logic (and meaning).

We can picture these distinct senses of possibility using the diagram in Figure 10.1, since the propositions that are nomologically possible are a subset of those that are logically possible. Anything that is possible in any sense is at least logically possible. For this reason, it is often referred to as 'possibility in the widest sense of the term.'

### EXERCISE 10.1

#### Nomological and Logical Possibility

For each of the following propositions, decide whether it is nomologically possible, logically possible, both, or neither. In a sentence, explain why.

- A. Jeff Goldblum had a dinosaur for breakfast.
- B. There is an object that weighs exactly 5 grams and exactly 7 grams at the same time.
- C. There is an electron that is both positively charged and negatively charged at the same time.
- D. There exist unicorns.
- E. There exists a cat that has plastic fur.
- F. There is a mind that exists without a body.

**De dicto modality:** the topic of the status of propositions (or dictums) as possible, impossible, necessary, or contingent

**Essential property:** a property that holds of an entity with necessity and makes it the kind of entity that it is

**De re modality:** the topic of whether individual entities have properties that are essential to them or merely contingent

So far, we have been talking about the modal status of propositions. Some are necessary, some contingent; some possible, some impossible. This sort of modality is what is called **de dicto modality**. Here we are discussing the status of a proposition or a dictum, hence 'dicto.' But we could also ask about the modal features of objects themselves. We could ask, for example, whether a certain basketball player has the possibility of being traded to another team. Or we could ask whether Venus is necessarily a planet. A historically interesting position in metaphysics is that objects have certain properties that hold of them necessarily, so-called **essential properties**. For example, if you are a Platonist and believe in entities like numbers, it is natural to think that the number 3 has certain essential features: it is necessarily prime, it is necessarily odd. It also has certain contingent features, such as numbering the moons of Pluto. More controversial is the issue of whether material objects like tables, chairs, or organisms have essential features. Was it only a contingent feature of Socrates that he was wise? Or was Socrates essentially wise? Are penguins essentially birds? Are persons essentially rational, thinking creatures, or could there be a human who was not rational? We'll discuss these issues later in the chapter. For now, we will just distinguish the issues of de dicto modality, concerning which propositions are necessary or contingent, possible or impossible, from the issues of **de re modality**, concerning which properties, if any, are had essentially or contingently or possibly by which objects. This is called 'de re' modality from the Latin for 'thing': res.

### 10.3 The Possible Worlds Analysis of Modality

We will begin by trying to better understand the phenomenon of de dicto modality, which is to some extent less controversial. Then in the last sections of this chapter, we will return to the issue of de re modality.

Let's start with a simple example of a modal truth to see where the main issues in the metaphysics of modality lie. There is a common expression used to discuss something that is unlikely to happen, 'when pigs fly.' Now pigs don't actually fly, but it's possible that they could. They might fly if they had wings. And there is no contradiction contained in the very idea of a pig with wings. That is, the proposition expressed by the following sentence is possibly true (in the sense of logical possibility):

(1) Pigs fly.

And its being possible means the following is just plain true:

(2) It is possible that pigs fly.

The central question that bothers metaphysicians is what could make a sentence like (2) true. Is there any way to understand (2) in terms that do not involve modal notions? Or does (2) just express a brute fact?

To take another example, we have seen that there are certain features of reality that couldn't be different than they actually are. For example, it is necessary that triangles have three sides. In other words, the proposition expressed by the following sentence seems necessarily true:

(3) Triangles have three sides.

And in virtue of its being necessarily true, the following is just plain true:

(4) Necessarily, triangles have three sides.

But again, what is the content of the 'necessarily' in this sentence? Is there a way to explain the truth of (4) in non-modal terms?

We want to talk about the metaphysics of modality. Following the methodology we have used in this book, it should be no surprise that most metaphysicians first approach the topic of modality by looking at the correct way to symbolize modal truths in first-order logic. There is a branch of logic that was developed in the early twentieth century to represent modal claims and assess the validity of arguments involving them: **modal logic**. Modal logic introduces new sentential operators to represent the possibility or necessity of individual propositions.<sup>2</sup> It was initially developed in the 1920s and 1930s by the philosopher and logician C.I. Lewis (1883-1964) as a form of propositional logic. In the 1940s, Ruth Barcan Marcus (1921-2012) and later Saul Kripke (1940-2022) extended Lewis's modal logic to include first-order quantification. Today, the most commonly used notation is the diamond,  $\Diamond$ , for representing possibility and the box,  $\Box$ , for representing necessity. So, for example, we might write:

$\Diamond$  (Pigs fly)

or

$\Diamond \exists x (Px \wedge Fx)$ ,

and

$\Box$  (Triangles have three sides)

or

$\Box \forall x (Tx \supset Sx)$ .

If you take a course in modal logic, you will learn new rules that will allow you to prove the validity of arguments using the  $\Box$  and  $\Diamond$ . You will learn that there is an equivalence between  $\Box A$  and  $\neg \Diamond \neg A$  for any sentence  $A$ . You will also learn that even in the weakest systems of modal logic, if  $A$  is a theorem in that system, that is, if  $A$  is a logical truth, then so is  $\Box A$ .

**Modal logic:** the branch of logic that deals with modal claims

**EXERCISE 10.2****Symbolizing Modal Claims Using the Box and the Diamond**

Using the key below, symbolize the following sentences using the notation of modal logic. Remember to bind all variables to quantifier expressions.

Mx: x is married

Bx: x is a bachelor

Dx: x drinks cocktails

- A. Married bachelors are impossible.
- B. Necessarily, all bachelors are unmarried.
- C. There exists someone who is a bachelor, and is necessarily unmarried.
- D. It is possible that there exists a bachelor who drinks cocktails.
- E. It is a contingent fact that some bachelors drink cocktails.

We won't worry about proving anything in modal logic here. Instead, what we will do is try to better understand what makes sentences using these modal operators true or false. In plain first-order predicate logic, we understood the truth conditions for our sentences to be given rather straightforwardly. For example, in a language where the name 'a' is assigned to Alice and 'Fx' stands for the predicate 'x is friendly,' we could understand what it would be for 'Fa' or ' $\exists x \neg Fx$ ' to be true. 'Fa' is true just in case Alice is friendly. ' $\exists x \neg Fx$ ' is true just in case there exists at least one person who is not friendly. But what does the world have to be like for ' $\Diamond \exists x Fx$ ' to be true? If Alice is friendly, this would seem to be enough to make ' $\Diamond \exists x Fx$ ' true. If Alice is friendly, then someone is friendly, and so it must be at least possible that someone is friendly. (What is actual can't be impossible.) But what if it turned out that no person who actually existed was friendly? It still seems the modal claim ' $\Diamond \exists x Fx$ ' would be true. There isn't any contradiction after all in the very idea of a friendly person. So, we need to have a more general understanding of what it takes to make modal claims asserting possibility true, one that could make ' $\Diamond \exists x Fx$ ' true even if no one was actually friendly.

One common way of understanding modal claims today is what is called the **possible worlds analysis of modality**. This view has historical origins tracing back to the work of Leibniz. It can be boiled down to two claims:

' $\Diamond A$ ' is true iff there is a possible world in which 'A' is true.<sup>3</sup>

' $\Box A$ ' is true iff 'A' is true at all possible worlds.

**Possible worlds**

**analysis of modality:** an analysis of claims about possibility and necessity in terms of what is true at various possible worlds (including the actual world)

So, according to the possible worlds analysis of modality, 'It is possible that someone is friendly' will be true just in case there is at least one possible world in which someone is friendly. There doesn't have to be any actually existing friendly person, just some friendly person at some possible world.

Similarly, according to the possible worlds analysis of modality, 'it's necessary that there are no round squares' is true because 'there are no round squares' is true at all possible worlds. There is no possible world that has a round square.

The possible worlds analysis is a very common way that logicians understand modal claims today. However, to a metaphysician, this analysis should immediately raise some eyebrows. What are these possible worlds? Are these worlds supposed to be literally existing alternative universes? If so, by saying the possible worlds analysis of modality is extremely common, am I saying that just by accepting the truth of 'Possibly, someone is friendly,' most philosophers thereby commit themselves to the existence of alternative universes?

The short answer is "it depends what you mean by 'universe.'" Some metaphysicians think they have a special way of understanding possible worlds so that they aren't literal universes. This is tricky and controversial, but first let's consider what has for the past several decades been the most provocative and hotly debated view about modality: David Lewis's view that the possible worlds appealed to in the possible worlds analysis of modality are literal universes, just as real as our own universe. This view is called **modal realism**.

Lewis defended this view most famously in his 1986 book *On the Plurality of Worlds*. There, he defined a world as a "maximally connected space-time region." Anything that is spatiotemporally connected to us is a part of our world. And so, when he uses the word 'world,' he doesn't mean something as small as our own planet Earth, but instead everything that exists at any spacetime distance from us. This will include all of the other planets, and indeed the rest of what we usually call our universe.

Each other world lies in its own spacetime. Worlds for him are concrete<sup>4</sup> and exist in the same sense as our world. Cars, planets, and people at these other worlds are no less real than those at our own world. They aren't ideas in our minds or fictions. For this reason, to say that our world is actual and the others are merely possible is just to say that the parts of those other worlds are not contained in our own spacetime. For Lewis, 'actual' is an indexical term like 'I' or 'here': what is actual for one depends on one's own perspective. It depends on what is the case in one's own spacetime.

There is one thing that is certainly appealing about Lewis's modal realism. This is how clear an analysis of modality it provides. If you want to know what makes it true that it is possible that pigs fly, his answer is

**Modal realism:** the view that in addition to the actual world, there exist other alternative universes, possible worlds, just as real as our own; and that it is in virtue of the nature of these universes that our modal claims are true or false

simple. This proposition is true because there literally are flying pigs at some other possible world, some spacetime disconnected from our own. And if you want to know what makes it true that necessarily triangles have three sides, again his answer is clear. It is because at every single world, every spacetime region, ours and all of the others, there never exists a triangle with anything other than three sides.

Despite its clarity, Lewis recognizes that his view will strike many as far-fetched. In a famous passage in *On the Plurality of Worlds*, Lewis notes the one objection to his view he cannot answer is the “incredulous stare.” Nonetheless he insists that however radical this view may sound, he has excellent reasons to believe it. This is because of the vast amount of theoretical work his concrete possible worlds are able to do. Let’s say a little bit about that.

Lewis calls his plurality of worlds a “philosophers’ paradise.” It is a philosophers’ paradise because, he notes, once one believes in the genuine existence of alternative possible worlds, one is able to exploit them in analyses that help us understand scores of otherwise perplexing phenomena. The big three phenomena on which we will focus are: (1) modality, (2) properties, and (3) content. However, this is really the tip of the iceberg, because once one has an account of these three, lots of other phenomena that have been interesting and puzzling to philosophers are much easier to understand.

Let’s start with modality. We have already seen that modal realism gives us a way to adopt the possible worlds analysis of modality and interpret it literally, thus demystifying modal talk. A modal realist is able to understand the box and diamond of modal logic as disguised quantifiers, quantifying over possible universes. If we let our variable  $w$  range over possible worlds, then:

‘ $\Diamond A$ ’ may be understood in the same way as:  $\exists w$  ( $A$  is true at  $w$ ).

‘ $\Box A$ ’ may be understood in the same way as:  $\forall w$  ( $A$  is true at  $w$ ).

But there are several other modal notions that belief in possible worlds can help one understand as well. For example, we think that some propositions are contingent; in other words, it is possible that they are true and also possible that they are not true. For example, we think it is a contingent fact that Manchester United wears red uniforms. It is possible (because it is actual) that they wear red uniforms. But it is also possible that they didn’t wear red, but uniforms of a different color, say, green. Contingency can now be understood as:

$A$  is contingent just in case  $\exists w$  ( $A$  is true at  $w$ ) and  $\exists w$  ( $A$  is not true at  $w$ ).

There are also special sorts of claims, counterfactuals, that Lewis’s analysis can help us better understand.

Counterfactuals are conditional claims that assert what would have been the case had things been different. For example, consider:

If the World Trade Center attack hadn't happened, then airport security would have been weaker.

We can use the symbol ' $\Box \rightarrow$ ' to represent this sort of conditional, where:

' $A \Box \rightarrow B$ ' may be read as: If A were the case, then B would be the case.

Then, supposing A did occur, the **counterfactual** ' $\neg A \Box \rightarrow B$ ' may be read as:

If A hadn't happened, then B would be the case.

Lewis (1973a), following Robert Stalnaker (1968), proposed analyses of counterfactual conditionals that appealed to what is true at the worlds that are most similar to the actual world. For example:

' $\neg A \Box \rightarrow B$ ' is true iff in all of the most similar possible worlds to the actual world where A doesn't occur, B occurs.

What Lewis and Stalnaker mean by a 'similar' world is a world that as much as possible contains the same kinds of objects instantiating the same kinds of properties and relations (while making the changes required for A to be false at that world). So, 'If the World Trade Center attack hadn't happened, airport security would have been weaker,' will be true on this analysis just in case in all of the worlds most similar to the actual world, but where the World Trade Center attack never occurs, airport security is weaker.

Similarly, consider one of Lewis's favorite examples:

If kangaroos had no tails, they'd topple over.

If we assume modal realism and the Lewis-Stalnaker analysis of counterfactuals, then for this to be true, it must be the case that in all of the other worlds most similar to our own but where kangaroos don't have tails, the kangaroos topple over.

There are lots of other modal notions we won't have time to discuss here that also get clearly illuminated if one is able to assume the existence of possible worlds: dispositions, supervenience, and physicalism, to name a few. One can find much more discussion in chapter 1 of *On the Plurality of Worlds* (Lewis 1986).

Another topic that is illuminated if we are permitted to quantify over other possible worlds is that of properties. In Chapter 2, we discussed class nominalism. This is the view that a property is just the class of its instances. For example, redness is just the class of red things, wisdom is the class of wise things, and so on. This is a view that will be appealing to those metaphysicians who (perhaps due to the indispensability argument) want to believe in classes, but for one reason or another find the existence of universals hard to swallow.

**Counterfactual:** a conditional asserting what would have been the case had things gone differently than how we suppose they actually go



In that earlier chapter, we noted a problem for class nominalism: the objection from coextension. Recall, this was the problem that there seemed to be cases in which two distinct properties have the same extension. This would be the case if, for example, the class of red things turned out to be identical to the class of round things. Even if this were so, we wouldn't want to say that the property of being red and the property of being round were the same property. But if one says a property just is the class of its instances, then in this case redness and roundness would turn out to be the same property. This is a place where Lewis's modal realism can help. If we believe in the existence of objects beyond those in our actual world, then we can distinguish these properties. For even if it turns out that all of the red things are actually round, it is possible that there are some red things that aren't round. And this just means for Lewis that there are other possible worlds where there are red things that aren't round things. The class nominalist may then identify properties with the set of all of the objects that actually and possibly instantiate them. Thus, our two properties are distinguished. The objection from coextension thus seems to be solved for the class nominalist. This response seems unavailable, however, if one does not believe in merely possible objects – objects that exist only at other possible worlds.

If we believe in possible worlds, we can also give an interesting account of the content of our thoughts and language. For our purposes here, let's just focus on the content of thoughts. Suppose at a given time *t* you are ignorant about a certain topic. Say you would like to know but don't yet know in what year the nation of Italy was founded. At this point you believe there is such a nation (Italy) and that it was founded in a certain year, but you don't have any belief about which year in particular it was founded. Here we may describe your state of mind in the following way. You believe you exist in one of the set of possible worlds in which a nation of Italy exists. But you can't narrow down the world you live in any further.

Now suppose you decide to find out when Italy was established as a nation. You look online and find out that Italy was founded in 1861. What this lets you do, one might say, is narrow in more closely on the set of worlds that could be actual according to what you believe. More generally, as one comes to form more and more beliefs, what one does is close in more and more narrowly on the set of worlds that could be the actual world.

Since Lewis is a realist about possible worlds, he can say that the content of one's belief literally is the set of possible worlds at which that belief is true. The content of the belief that Italy was founded in 1861 is the set of worlds at which Italy was founded in 1861. This belief is true just in case the actual world is a member of this set of worlds. The belief is false just in case the actual world is not a member of this set of worlds. So, in addition

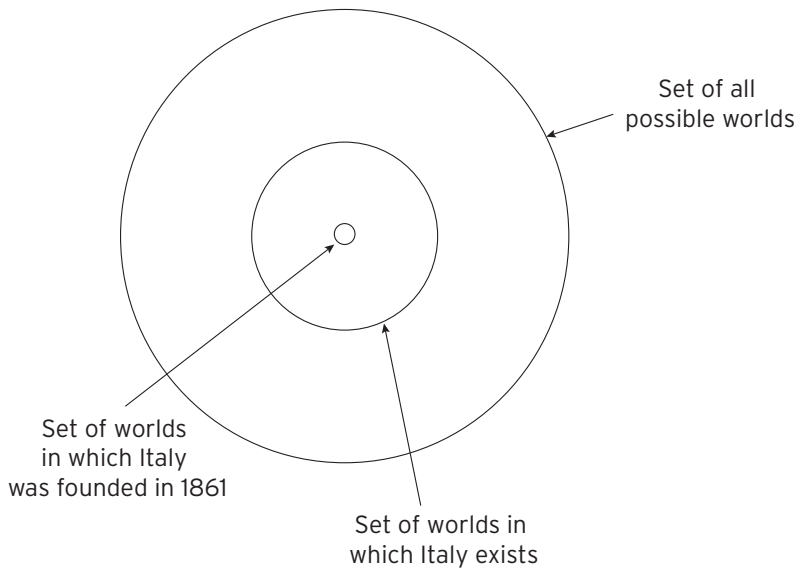


Figure 10.2 The Content of Beliefs as Sets of Possible Worlds

to giving us a way to understand (1) modality and (2) the nature of properties, modal realism also gives one (3) an intuitive and clear analysis of mental content.

### EXERCISE 10.3

#### The Content of Beliefs as Sets of Possible Worlds

The theory that the content of a belief is the set of possible worlds at which it is true is initially quite intuitive. It makes sense to many to think that what one is doing as one forms more beliefs is closing in more narrowly on the set of worlds one takes to be the actual world. But there are some difficulties for this view. For example, many take mathematical truths to be necessary. It is, for example, a necessary truth that  $\pi$  is an irrational number. There is no possible world in which  $\pi$  is rational.<sup>5</sup> Why would this present a problem for the view that the content of a belief is the set of possible worlds at which it is true? What could a modal realist say to allay the concern?

As I have mentioned, these are only some of the useful theoretical applications of modal realism. Of course, it is a bold and controversial claim that there really exist all of these universes distinct from our own, but what Lewis argues is that the hypothesis of the existence of these worlds lets us do so much explanatory work (in understanding modality, properties, content, and more) that this gives us reason to believe the

hypothesis. Like Quine, whose views we considered in the first chapter, Lewis believed that a choice between ontologies should be settled in a way very much analogous to the way choices are made between scientific theories. The metaphysician, like the scientist, postulates a realm of what-ever objects are necessary to achieve explanations that are simple and powerful; he or she seeks out the hypotheses that can explain the most in the fewest possible terms. Lewis makes a good case that just by positing other possible worlds, he can explain an extremely wide range of perplexing theoretical phenomena.

Still, unsurprisingly, most philosophers have not been convinced by Lewis that there really are such things as these other possible universes. And this raises the question of whether there may be an alternative way to understand possible worlds that dials back on the ontological commitments while at the same time manages to achieve all of the explanatory power of Lewis's framework. One strategy is to postulate possible worlds while conceiving them as something other than additional concrete universes. Those who adopt this type of view are called **ersatz modal realists** ('ersatz' means fake). This position, and the challenges it faces, will be the topic of the next section.

**Ersatz modal realism:**

the view that there are possible worlds (worlds that play a similar role to the concrete worlds of the modal realist), but that these are not additional universes in the same sense as our universe

**TEXT BOX 10.2**

**Modal Realism and the Multiverse**

If you are familiar with developments in recent cosmology, you'll notice that Lewis's proposal that there exist a plurality of possible worlds, spacetimes disconnected from our own with inhabitants just as real as us, bears strong similarities to the sort of multiverse hypotheses defended today in cosmology. Indeed, many think that the most promising account of our early universe, the inflationary hypothesis, straightforwardly entails that there are many universes disconnected from us so that we can never observe them.

There is certainly a similarity between modal realism and the multiverse hypothesis. Some, like the metaphysician Alistair Wilson (2020), have argued that we can identify Lewisian possible worlds with the worlds in the physicist's multiverse. This is quite interesting given that the reasons Lewis adopted for his hypothesis are philosophical and, many would argue, purely a priori, while the reasons cosmologists have for endorsing their hypothesis are scientific, and many would argue, largely empirical. Still there are real differences between the philosophical proposal of Lewis and these scientific proposals. First, although the alternative universes of contemporary cosmology appear *causally* isolated from one another, they are not genuinely *spatiotemporally* disconnected in the way Lewis's universes are. Moreover, it seems clear that the cosmologist is proposing these extra universes not as alternative possibilities, or ways things might have been, but as additions to how things actually are. These are not mere possibilities according to the inflationary cosmologist, but more parts of actuality. Thus, it is not clear Lewis would be satisfied with them as part of his analysis of modality.

## 10.4 Ersatz Modal Realism

There are many different forms of ersatz modal realism. Alternative versions of the view have been proposed by Peter Forrest (1986), Alvin Plantinga (1992), Robert Stalnaker (1976), and many others. Ersatz modal realists typically adopt a common strategy; accept the possible worlds analysis of modality (and of counterfactuals, properties, content, and so on) but don't think of the worlds figuring in these analyses literally as alternative concrete universes. Instead, think of these worlds as objects constructed out of things we already believe in as existing at the actual world. The ersatz modal realist will typically construe his or her possible worlds as abstract rather than as concrete entities. In this section, we will focus on one very natural way of thinking of these worlds, as linguistic items, sentences (very long sentences) that describe alternative ways things might have been. To find alternative ways philosophers have thought of these ersatz worlds, consult the suggestions for further reading at the end of this chapter.

Because ersatz modal realists construct possible worlds out of things that actually exist, they don't have to believe in anything that is *merely* possible. Ersatz modal realism is thus one way to hold on to **actualism**: the position that everything that exists actually exists. (We will discuss other ways to be an actualist while rejecting ersatz modal realism in the next section.) Actualism contrasts with **possibilism**, the view that there are at least some entities that are merely possible. This is Lewis's position of course, since he believes there are some entities that are not actual: all of those things that exist in the spacetimes disconnected from our own.

The kind of **linguistic ersatzism** (short for 'linguistic ersatz modal realism') we'll discuss here is criticized by Lewis in his book and we'll get to that criticism in a moment, but it is worth pointing out first how intuitive the view is. It is not just a way of getting everything Lewis gets without having to buy into the ontological commitments of modal realism; it is also a very natural position on its own. You might ask yourself a second: What does it mean to say that a certain proposition is possible or impossible? For example, what does it mean to say it is possible that there are talking donkeys or impossible that there are round squares? One natural response to these questions is that the former is possible because if you said 'There are talking donkeys' you wouldn't be contradicting yourself despite saying something false, while if you said 'There is a round square,' you would be saying something that involves a contradiction.

Linguistic ersatzism captures this intuition by conceiving of possible worlds as maximal consistent sentences, long conjunctive sentences of the form ' $A \wedge B \wedge C \wedge D \wedge \dots$ ' that are maximal in the sense that for any basic or atomic sentence ( $A$ ,  $B$ ,  $C$ , etc.) in the language in which these sentences are constructed, either it or its negation is contained as a conjunct in that sentence. According to the linguistic ersatzer, a proposition  $p$  is

**Actualism:** the view that everything that exists actually exists

**Possibilism:** the view that there are at least some entities that are merely possible

**Linguistic ersatzism:** a form of ersatz modal realism that takes possible worlds to be sentences or other linguistic entities

possibly true just in case it is true according to some possible world, that is, iff some maximal, consistent sentence in the ersatzers's worldmaking language entails that *p*.

*The Linguistic Ersatzers's Way of Understanding the Possible Worlds Analysis of Modality*

' $\Diamond A$ ' is true iff some maximal consistent sentence in the ersatzers's worldmaking language entails that *A*.

' $\Box A$ ' is true iff every maximal consistent sentence in the ersatzers's worldmaking language entails that *A*.

It is up to the linguistic ersatzer (a) how she wants to think of these sentences, as abstract or concrete, and (b) which language she wants to use to build up her worlds. One thing that will be important is that she has a rich enough language to be able to build up sentences describing all of the possible ways a world might be. An initial obvious problem for the linguistic ersatzer is that she isn't going to be able to use a natural language like English or Japanese to construct her worlds since no natural language has a name for every actual object there is, let alone every possible object there could be.

What we will discuss here is the problem for linguistic ersatzism that Lewis finds most pressing and this is that the view does not, unlike his own view, really provide an analysis of possibility in terms of antecedently understood notions like existence and spacetime, but instead presupposes an understanding of modality. If this is so, the view cannot solve the problem the metaphysician of modality is aiming at: to explicate what we mean when we say something is possible or impossible, necessary or contingent. To see how linguistic ersatzism fails to analyze, but instead presupposes an understanding of possibility, Lewis presents the ersatzer with a dilemma. When she chooses a language in terms of which she will construct her possible worlds, the linguistic ersatzer faces the following choice. On the one hand, she may use what Lewis calls a rich worldmaking language. On the other hand, she may use what Lewis calls a poor worldmaking language. Either way, the ersatzer faces a problem. Let's start with the first option.

### TEXT BOX 10.3

#### Finding a Worldmaking Language

To find a worldmaking language with enough terms to capture all of what is possible, Lewis suggests that the ersatzer use a language in which entities function as names for themselves. In *Gulliver's Travels*, Gulliver visits the town of Lagado, where the residents use objects as names for themselves. So, Lewis suggests calling this a 'Lagadonian language.' Using a Lagadonian language, at least the ersatzer will have a name for every actual object or property. This will be enough for some ersatzers. Others may think we need names for not just all of the actual entities but also all of the merely possible entities there could be.<sup>a</sup> This would require more resources than a Lagadonian language has to offer.

To say that the ersatzer is using a rich worldmaking language is to say that their worlds are built up using a language rich enough to state all of the very diverse kinds of possibilities we think there are simply and directly. For example, consider the following sentences:

- (1) There are talking donkeys.
- (2) The COVID-19 pandemic did not occur.
- (3) There are unicorns.
- (4) Barack Obama is a world-class tennis player.

Suppose these sentences are all actually false, but possibly true. Then, according to the ersatzer, for each of these sentences, there must be a possible world, a maximal consistent sentence in her worldmaking language, that entails it is true. If they are using a worldmaking language that is rich, that has enough structure to state these possibilities simply and directly, then there will be maximal consistent sentences in their language with parts that say each of (1) - (4). So, there will be a part of one of these sentences that says that there is a talking donkey, a part of another that says Barack Obama is a world-class tennis player, and so on.

So far there is no problem. The trouble comes when we ask what determines which possible worlds there are and which there aren't - what determines which are the maximal *consistent* sentences? Recall the linguistic ersatzer's analysis:

' $\Diamond A$ ' is true iff some maximal *consistent* sentence in the ersatzer's worldmaking language entails that A.

' $\Box A$ ' is true iff every maximal *consistent* sentence in the ersatzer's worldmaking language entail that A.

To see the worry, we might ask, is there a possible world, according to our ersatzer, that entails the following:

- (O) Barack Obama is a world-class tennis player and Barack Obama has never picked up a tennis racket.

The answer to our question should be "no." Being a world-class tennis player requires picking up a tennis racket and knowing how to swing it. So, no maximal *consistent* sentence should entail (O). (O) intuitively involves a contradiction.

However, it is going to be challenging for the ersatzer to give an account of what makes this sentence inconsistent in a way that doesn't simply appeal to the notion of possibility. If there were a logical contradiction in this sentence, something of the form  $[A \wedge \neg A]$ , then the ersatzer could simply apply a rule that nothing with that form ever appear in their language. The trouble is that even though it is impossible to become a world-class tennis player while never picking up a racket, this is not because there is a logical contradiction contained in this thought. Yet if the ersatzer just builds into their account the claim that the maximal consistent sentences are the ones that are possible, then they will be unsuccessful in their

explanation of what it means to say something is possible. The explanation they provide will be circular.

There is a response the ersatzers can make here. They might propose that all of their ersatz worlds also contain a set of what we might term the 'metaphysical truths,' truths such that if we build them into our maximal sentences, they will allow us to eliminate contradictions in a straightforward way because they will entail logical inconsistencies. For example, one might propose, 'For all  $x$ , if  $x$  is a world-class tennis player, then  $x$  has picked up a tennis racket' as one of these metaphysical truths. This would then entail that no maximal consistent sentence in their language will ever say (O). However, there is a question whether this project of discovering all of the metaphysical truths can be carried out. In any case, we seem to have postponed the project of analyzing modality until this is done.

This brings us to the other horn of Lewis's dilemma. The linguistic ersatzer also has the option of constructing their ersatz worlds using what Lewis calls a poor worldmaking language. This would be a language consisting of a small set of basic terms that are sufficient to construct all of the possibilities there are. Here the hope would be that the ersatzer can allow for the sentences we considered above to come out as possibly true. ('There are talking donkeys,' 'The COVID-19 pandemic never occurred,' etc.) But this will not be done by using a worldmaking language in which such claims are expressed simply and directly. The motivation for using a poorer worldmaking language is to avoid the problem we just considered: to give a straightforward account of which sentences in the worldmaking language are consistent and which are inconsistent without circularity. The move to a poor worldmaking language is an attempt to give the ersatzer a shot at a reductive analysis of modality, an analysis that gives an account of all modal claims in nonmodal terms.

There are several ways in which one might try to set up a poor worldmaking language. One would be to start with a list of names, ' $a_1$ ,' ' $a_2$ ,' and so on, that refer to all of the fundamental objects in the world (perhaps the elementary particles) and then a list of predicates, ' $F_1$ ,' ' $F_2$ ,' and so on, that refer to all of the fundamental properties and relations there are. Then, each sentence (each ersatz world) will be a conjunction of basic sentences saying for each fundamental object whether for each property it has that property or not.

Another kind of poor worldmaking language that Lewis considers uses the real numbers to denote locations in spacetime. Ersatz worlds are then constructed out of sets of ordered pairs in which the first member is a number naming a spacetime location and the second member is a zero or a one depending on whether that location is occupied with matter or

not. Here the sentences in the worldmaking language are sets. A maximal world sentence will have members corresponding to each spacetime location in that world. So, for example, an ersatz world in this language might be written like:  $\{ \langle 0.0001, 0 \rangle, \langle 0.0002, 0 \rangle, \dots, \langle 1.00001, 1 \rangle, \langle 1.0002, 1 \rangle, \langle 1.0003, 1 \rangle, \dots \}$ .

The advantage of using a poor worldmaking language is that an ersatzter who constructs their worlds in this way does not have to appeal to our pre-theoretic grasp on what is and is not possible in order to say which worlds, which maximal sentences, are consistent and which are not. Depending on their choice of worldmaking language, the ersatzter can then give a simple and straightforward account of what consistency amounts to. In the first language, they can say a consistent sentence is one in which for each name  $a_i$  and predicate  $F_i$ , either  $F_i a_i$  or  $\neg F_i a_i$  appears as a conjunct in the sentence, but not both. Here, consistency reduces to logical consistency. In the second case of a poor worldmaking language we considered, a consistent sentence will be one in which no number denoting a spacetime location is paired to both a 0 and a 1. This account again is straightforward and does not presuppose any antecedent grasp of what is and is not possible.

However, there is a different kind of problem for the linguistic ersatzter who appeals to a poor worldmaking language. Recall, the linguistic ersatzter wants to say:

' $\Diamond A$ ' is true iff some maximal consistent sentence in the ersatzter's worldmaking language entails that A.

' $\Box A$ ' is true iff every maximal consistent sentence in the ersatzter's worldmaking language entails that A.

But if this is right, and we use one of the above poor worldmaking languages, then none of the facts we considered above will turn out to be possible. Recall, these are:

- (1) There are talking donkeys.
- (2) The COVID-19 pandemic did not occur.
- (3) There are unicorns.
- (4) Barack Obama is a world-class tennis player.

At least in neither case of the poor worldmaking languages we considered above will there be ersatz worlds that entail any of these sentences. The worlds will describe fundamental particles and their properties or points of spacetime and whether they are occupied or unoccupied. But they will say nothing about donkeys or unicorns or tennis players.

Intuitively, what the ersatzter who likes a poor worldmaking language will want to say is that even though their worlds don't in so many words say that there are unicorns or talking donkeys, still some of these worlds



are such as to *entail* facts about the existence of unicorns or talking donkeys, the ones that describe the correct arrangements of fundamental properties instantiated by fundamental objects or the correct patterns of filled spacetime. But now a modal realist will press on this notion of entailment. Entailment is itself a modal notion. To say a certain pattern of filled spacetime entails the existence of a talking donkey means that *there couldn't be* such a pattern of filled spacetime without the existence of a talking donkey. And so, the linguistic ersatzers who use a poor world-making language to construct their worlds also seems to be presupposing modality in their account.

As with the case of the rich worldmaking language, there is a reply the ersatzers might make. They might hope to be able to provide a list of facts about which patterns of fundamental stuff correspond to which nonfundamental possibilities and just append this list of facts to their analysis. (E.g., such and such pattern of occupied spacetime corresponds to a talking donkey, thus and so pattern of occupied spacetime corresponds to a unicorn, and so on.) However, again, this will be a substantial undertaking. It is questionable whether anything like this can be carried out, and even if it can, it only postpones the analysis of modality.

Thus, although an ersatz version of modal realism is a live option, such a position does have well-known difficulties. The question of how best to overcome them makes the view much less clear and worked-out than Lewis's own position.

## 10.5 Rejecting the Possible Worlds Analysis

The possible worlds analysis of modality is by now deeply entrenched in metaphysics, as well as philosophy more generally. This analysis finds wide use in the philosophy of language, epistemology, ethics, the philosophy of science, and other areas, as philosophers often have reason to appeal to modal notions in their work. Philosophers across these sub-disciplines will often slip back and forth unconsciously between talk of what is possible and what exists in some or other possible world, what must be and what is the case in all possible worlds. But we have now seen there are deep metaphysical questions that arise with appeals to possible worlds language. One must either adopt Lewis's modal realism and buy into a controversial ontology of many alternative universes or find some alternative and yet clear way of understanding these possible worlds as constructed out of sentences or other actual entities.

The problems with both alternatives have caused some metaphysicians to insist that all of this possible worlds talk, convenient as it is, must not be the correct way to analyze our modal language. It may be useful, evocative as a *façon de parler*, but it shouldn't be taken as anything more than

that, a way of speaking. To find out what we really mean when we say that such and such is possible or necessary, we need to pursue an alternative analysis.

One interesting contemporary alternative to the possible worlds analysis of modality is modal **fictionalism**. We have already encountered fictionalism in Chapter 8.<sup>6</sup> Here again, fictionalism is a way of understanding the truth conditions of a certain domain of claims that appear to carry ontological commitments, while avoiding these ontological commitments. We will consider one version: the modal fictionalism of Gideon Rosen (1990).

The primary motivation for the fictionalist is to avoid ontological commitment to alternative possible worlds. According to the fictionalist, there are no concrete alternative universes like Lewis thought. And there are no ersatz worlds either. So, the following must be rejected:

*The Possible Worlds Analysis of Modality*

A is possible just in case there is a possible world in which A is true.

A is necessary just in case A is true at every possible world.

Still, the fictionalist believes that there are lots of true modal claims, claims about what is possible and what is necessary. And the fictionalist does accept possible worlds talk as an evocative and useful way of thinking about what is possible and necessary. Their proposal is that we can use possible worlds talk, but that we should replace the Possible Worlds Analysis of Modality with the following fictionalist analysis of modality:

*The Fictionalist Analysis of Modality*

A is possible just in case according to the fiction that there are possible worlds, there is a possible world in which A is true.

A is necessary just in case according to the fiction that there are possible worlds, A is true at all possible worlds.

Note that for there to be modal truths on this analysis, no possible worlds need exist, only fictions, stories, about the existence of possible worlds. And of course no one questions the existence of such stories. One such fiction, as Rosen cheekily points out, is Lewis's own *On the Plurality of Worlds*.

Many questions have been raised about the adequacy of modal fictionalism to account for our modal claims. Among the questions some have raised is whether the fictions about possible worlds that exist (including Lewis's book) really say enough to give a complete account of all of the modal truths that there are. It is also frequently asked whether it follows from Rosen's account that there were no facts about what was possible or necessary before the publication of Lewis's book in 1986. (He says it doesn't follow.)

**Fictionalism:** the view that truths in a certain domain should be understood as truths about a fiction

**EXERCISE 10.4****Rejecting Commitment to Possible Worlds**

Why do you suppose Rosen and other philosophers have been skeptical about the existence of possible worlds? What reasons are there for a philosopher to be skeptical about the existence of ersatz possible worlds? Are these good reasons to reject modal realism and ersatz modal realism? Which view do you prefer?

**Conventionalism:** a view that seeks to reduce modal claims to claims about what follows or does not follow given the conventions of our language

Fictionalism is one strategy for understanding possibility and necessity while avoiding commitment to possible worlds (genuine or ersatz). Another strategy, one that before the 1970s and 1980s was dominant, was that alluded to in the beginning of this chapter: **conventionalism**. Conventionalists try to reduce modal claims to facts about what is or is not true according to the conventions of our language. A conventionalist account of modality completely avoids talk of possible worlds. Instead, those sentences that are necessary will be those that are entailed by the meanings of the terms constituting those sentences. For example:

Triangles have three sides.  
All bachelors are unmarried.

These are necessary truths because they are entailed by the meanings of the terms 'triangle,' 'bachelor,' 'unmarried,' etc. Similarly, the following sentences are contingent truths:

There are no pink donkeys.  
All members of the European Union lie north of the Equator.

Their contingency, according to the conventionalist, consists in the fact that the conventions of our language do not entail by themselves that these sentences are either true or false. It is consistent with the way we use the word 'donkey,' that there be no pink donkeys. It is also consistent with the conventions of our language that there be pink donkeys. The meaning of the word 'donkey' does not settle the matter of whether donkeys are pink or not pink. A sentence is contingent if its truth depends on what the world is like (beyond the facts about how we use our language). On the other hand, consider:

There exists a round square.

This is a sentence that is necessarily false or impossible. The conventionalist will explain this fact again by appealing to facts about how our

language works. Based on the meanings we give to the terms 'round' and 'square,' they will say that something's being a square rules out its being round. And so, this sentence is not just false, it is impossible.

The conventionalist's account of modality will be appealing to those who view ascriptions of possibility, necessity, or contingency to not be about what sort of things exist at our world and what they are like, but instead about our language, the kinds of things our language permits us to say or not say. If you just understand how our language works, you will in principle be able to figure out all of the necessary truths: what is possible, and what is impossible.

The reason why conventionalism is especially controversial today can be traced to some arguments developed in the 1970s by the philosophers Saul Kripke (1970/1980) and Hilary Putnam (1975). According to the conventionalist, necessary truths are supposed to be truths that are entailed simply by the meanings of our terms. But Kripke and Putnam noted that there were a class of necessary truths, the **necessary a posteriori**, that are not knowable through the meanings of the terms involved in them. These necessary truths had to be empirically discovered. Although these are necessary truths, what makes them true is not facts about our language, but facts about our world. Consider:

Gold has atomic number 79.

Water is H<sub>2</sub>O.

Both of these are necessary truths. Something cannot be gold if the atoms that make it up do not have 79 protons. This is something scientists had to learn about the chemical structure of gold. Similarly, a substance cannot be water if it does not consist of molecules containing two hydrogen atoms and an oxygen atom. But these facts aren't contained in the meanings of the words 'gold' and 'water.' 'Gold' and 'water' are words that were used for centuries and could have continued to be used without our ever discovering modern chemistry. You do not have to understand facts about chemistry in order to understand the meaning of these terms and use them, to understand perfectly well that a certain ring is gold or a certain glass contains water. Given these examples, the necessary truths cannot simply be the truths that hold in virtue of the meanings of our terms and the conventions of our usage. Just by reflecting on the meanings of these terms and how we use them, we never could have discovered these modal facts.

The other kind of claim that threatens conventionalism is the **contingent a priori**. The conventionalist would like to say that contingent truths are those that are not made true merely by the conventions of our language, but depend on how the world turns out (independent of how we use our terms). But Kripke raised the following example of what

**Necessary a posteriori:**  
truths that are necessary,  
yet known on the basis of  
empirical observation

**Contingent a priori:**  
truths that are neither  
necessary nor impossible,  
and yet are discoverable  
merely by reflection on  
the meaning of the terms  
or concepts involved in  
them

appears to be a contingent claim that we could discover just by reflecting on the meaning of our terms:

The standard meter stick is one meter long.

The standard meter stick is an object kept in the International Bureau of Weights and Measures (BIPM) in Sèvres, France. This is an object that has been used to define what length it is that our term 'one meter long' refers to. For this reason, it is true just in virtue of the meaning of the terms involved that the standard meter stick is one meter long.

And yet, at the same time, it is a contingent fact about this meter stick that it is the length it is (one meter long). It could have been the case that the standard meter stick, that very object, wasn't used as the standard for the meter at all, but used for some other purpose. It could have been used to hammer a nail and broken in half in which case it wouldn't have been one meter long after all. It is a contingent fact that that object, the standard meter stick, is the length it is. And so, even though the above sentence is true in virtue of how we actually use the term 'one meter long,' it is not a necessary truth.

It is for these reasons that many philosophers conclude, although conventionalism at first seems to be a good idea, the facts about what is necessary and what is contingent do seem to concern (in at least some cases) something more than the facts about how we as a matter of fact use our terms. The conventionalist will need to find a way to argue against these cases if they want to reduce modality to facts about language.

#### **EXERCISE 10.5**

#### **The Necessary A Posteriori and Contingent A Priori**

What other examples of the necessary a posteriori and the contingent a priori can you think of? Provide an additional example of each.

### **10.6 Essentialism and Anti-Essentialism**

We can now finally discuss de re modality: what are the modal features, if any, of objects? Recall, this contrasts with the topic of de dicto modality, or which propositions are possible, necessary, or contingent. Here, we'll focus on the most central issue, which is the question of essentialism. We will use 'essentialism' to denote the view that objects themselves, independently of any ways we may have of thinking about or categorizing them, have essential properties. An essential property is a property of

an object such that were that object to fail to have that property, it would cease to exist. It is a property an object must have if it is to exist.

Essentialism is an old view in philosophy that traces back to Aristotle who talks in his *Metaphysics* about objects possessing essences, features it is in an object's nature to possess. The position survives into the modern period with Descartes who famously claimed that the essence of a mind is to think and the essence of a material body is to be extended in space. John Locke (1632-1704) as well discusses objects' *real* essences, those features that make an object the kind of object it is and explain its observable features.

However, by the middle of the twentieth century, with the influence of logical positivism and the transition into naturalistic metaphysics, philosophers became skeptical of essential properties. Quine's critique of the entire notion of de re modality is perhaps the most famous and so we will discuss this here and why he may be considered perhaps the most influential contemporary anti-essentialist.

In his book *Word and Object* (1960), Quine expresses puzzlement with the idea of de re modality. He asks us to consider the case of a certain individual (call him 'Jones') who is at the same time both a mathematician and a bicyclist. Surely, Quine suggests, the following claims are both true:

Mathematicians are necessarily rational and not necessarily two-legged.

Cyclists are necessarily two-legged and not necessarily rational.

And yet what does this mean for Jones and his essential features? Surely, Quine complains, "there is no semblance of sense in rating some of his attributes as necessary and others as contingent." (1960, p. 199)

Note that Quine doesn't seem to be particularly puzzled with the use of modal operators in general. Relative to a conceptual scheme or way of categorizing an object, it makes perfect sense to say an individual is essentially rational or essentially two-legged. The puzzlement comes when one asserts that objects may have certain essential properties not relative to a given classification, but in themselves, independently of how we think of them. We may describe Quine in this way as being particularly puzzled about de re modality and the commitment to essentialism.

We may more clearly see the distinction by considering two ways of regimenting a claim that if one is a mathematician, then necessarily, one is rational. Using the ' $\Box$ ' introduced in Section 10.2, ' $Mx$ ' for ' $x$  is a mathematician' and ' $Rx$ ' for ' $x$  is rational,' we may consider these two regimentations of 'Necessarily, mathematicians are rational':

(DD)  $\Box \forall x(Mx \supset Rx)$

(DR)  $\forall x(Mx \supset \Box Rx)$

The first (DD) is what metaphysicians think of as the *de dicto* reading. It says it is a necessary truth that all mathematicians are rational. To accept this, we need not think that there are any objects possessing essential properties, properties such that if the objects failed to have them, they couldn't exist. Rather we just need to think that it is a necessary truth that all mathematicians are rational. This could be true in virtue of the fact that at every possible world, if something is a mathematician, then it also happens to be rational.

On the other hand, the second claim, (DR), is what is thought of as the *de re* reading. We can see the difference between the *de dicto* and *de re* readings corresponds to a difference in the relative placement of the quantifiers and modal operators. In the *de re* reading, the quantifier takes wide scope outside of the modal operator. In the *de dicto* reading, the quantifier takes narrow scope and lies inside the modal operator. It is only the second *de re* reading that implies the instantiation (given the existence of mathematicians) of essential properties. For what (DR) says is that if a mathematician exists, then this mathematician has the property of being essentially rational. This means in any world in which this object *x* exists (whether in that world it is a mathematician or not), *x* has the property of being rational. If one is skeptical of essentialism, then one should also be skeptical about the truth of claims like (DR) in which quantifiers take wide scope outside of modal operators.

## 10.7 Essentialism Today

Essentialism remains a controversial position today, but not all philosophers want to reject it as Quine did. Indeed, in his *Naming and Necessity* (1970/1980), Kripke argued in defense of essentialism.

One of Kripke's most interesting claims in this regard is what is called **origins essentialism**. This is the claim that the origins of material objects, like tables and chairs, as well as those of organisms including human beings, are essential to them. For example, consider the Queen of England, Elizabeth II. Kripke asks his reader to consider whether it is plausible that this person, Queen Elizabeth, could have existed and yet had other parents than she actually had. Could it have been the case, for example, that Queen Elizabeth's parents were Mr. and Mrs. Harry Truman?<sup>7</sup> Kripke thinks not:

Can we imagine a situation in which it would have happened that this very woman the Queen came out of Mr. and Mrs. Truman? They might have had a child resembling her in many properties. Perhaps in some possible world Mr. and Mrs. Truman even had a child who actually became the Queen of England and was even passed off as

**Origins essentialism:**  
the view that the origins  
of material objects and  
organisms are essential  
to them

the child of other parents. This still would not be a situation in which *this very woman* whom we call 'Elizabeth II' was the child of Mr. and Mrs. Truman, or so it seems to me. It would be a situation in which there was some other woman who had many of the properties that are in fact true of Elizabeth.... It seems to me that anything coming from a different origin would not be this object. (Kripke 1970/1980, pp. 112-113)

Kripke extends this argument not just to human beings and their origins, but also to material objects as well. He asks us to think about a particular wooden table and conjectures we will agree with him that it is not possible to imagine a situation in which literally this same object exists and yet it has its origin in a different piece of wood. If the table were made of a different piece of wood, it would be a different table, not the same table. Thus, it has its origin essentially.

As we saw earlier, there are some philosophers, most notably Quine, who would be skeptical of claims like this one of Kripke's. Kripke takes it to be obvious that this very person, the Queen, or that very object, the table, could not have existed without having the same origins. But Quine wonders what could make a claim like this true. Take for instance the case of the Queen. Yes, Quine would say, if we consider this object the Queen in a particular way, relative to a classification as a particular organism say, then we may consider it to be an essential feature of hers that she is the daughter of these particular parents (George VI and Elizabeth, Duchess of York). This is just like the case of Jones. Relative to a particular way of thinking about him, as a mathematician, we are disposed to think of certain of his features as essential. But we may think of the Queen in another way, not worrying so much about her as an organism coming from a particular sperm and egg, but thinking of her simply as the English queen of more than eighty years. Then we might say it is not essential that this queen had these particular parents, but only that she had whatever particular features enabled her to reign as queen.<sup>8</sup>

It is common to find metaphysicians on both sides of the essentialist divide today. For some, Kripke is just simply right that independently of how we think about the Queen, she couldn't have been the very same person if she weren't born of the same parents. For others, Kripke's intuitions are better accounted for in terms of the way we think about these objects in certain contexts.

What is slightly less contentious is the claim that objects have other less robust sorts of essential properties. For example, one might less controversially adopt the view **sortal essentialism**, that it is essential to objects what kinds of things they are. So, the Queen couldn't exist without

**Sortal essentialism:** the view that it is essential to objects what kind of things they are



being a person. And this table, this very object, couldn't have existed without being a table.

Perhaps even less controversial is the claim that objects have their self-identity necessarily. It is true that necessarily every object is identical to itself. Does this mean that every object has at least the essential property of being self-identical? This certainly seems trivial, doesn't it? Or doesn't every object necessarily have the feature of being such that  $2+2 = 4$ ? Do claims like this then make essentialism trivially true? Ruth Barcan Marcus (1967) argued for a distinction between a class of "traditional" essentialist theses and other weaker essentialist theses that even the Quinean should accept. What is distinctive about the traditional essentialist theses is that they claim there are certain features that are necessary to some objects but not to all.

Even so, we should distinguish the *de dicto* and *de re* interpretations of these claims. Consider the claim that every object is necessarily self-identical. This may be symbolized as:

$$\Box \forall x(x = x),$$

or as:

$$\forall x \Box(x = x).$$

The first, *de dicto* claim, just says it is a necessary truth that every object is self-identical. It doesn't posit any essential features of individual objects as the second does. One may read the second, *de re* claim as saying: take any object you like, *that object* has the essential feature of being self-identical. A skeptic about *de re* modality can say that the trivial claim that everything is necessarily self-identical should be regimented using the *de dicto* claim. They can thus avoid commitment to any essential properties.

Finally, it is worth closing by mentioning that although it has been common in the past several decades to view the question of essentialism as the question of which properties are had necessarily by an object, if it is to exist, there is another tradition, another way of thinking about essence, that does not tie essence to necessity in this way.

Kit Fine argues for detaching the question of an object's essence from questions about necessity in his paper "Essence and Modality." He argues that it is better to think of an essence primarily as capturing what an object is in itself, its being. This will certainly have consequences for an object's necessarily being a certain way, but that it is necessarily some way doesn't make it what it is.

This may help us to see trivial properties such as being such that  $2+2 = 4$  as irrelevant to essentialism in any sense of the word. It is true this is a feature any object has necessarily. But this doesn't make it a

part of the object's essence, something that captures what that object is. According to Fine, essences should be thought of more as definitions, accounts that define what an object is.

## Suggestions for Further Reading

For Lewis and Stalnaker's theories of counterfactual conditionals, see Stalnaker's "A Theory of Conditionals" and Lewis's "Counterfactuals and Comparative Similarity." For more on the debate between the modal realist and the ersatz, see the papers by Phillip Bricker and Joseph Melia in *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne, and Dean Zimmerman. A widely influential version of ersatzism is described and defended in Alvin Plantinga's *The Nature of Necessity*. Others are defended by David Armstrong in his *A Combinatorial Theory of Possibility*, Peter Forrest's "Ways Worlds Could Be," and Robert Stalnaker's "Possible Worlds." An anthology containing many of the classic articles in the debate is *The Possible and the Actual*, edited by Michael Loux. For more on essentialism, see Richard Cartwright's "Some Remarks on Essentialism" and L.A. Paul's "In Defense of Essentialism." Marcus's "A Backward Look" paper gives a helpful overview of Quine's evolving distrust of de re modality as well as a critique. For discussion of modal epistemology, how we can know what is possible (or what is necessary), see the essays in Tamar Szabó Gendler and John Hawthorne's collection, *Conceivability and Possibility*.

## Notes

- 1 Sometimes the term 'logical possibility' is reserved for propositions that do not entail any contradictions on the basis of their logical form alone. In other words, something is logically possible if and only if, on the basis of logic alone, we could not deduce something of the form  $[A \wedge \neg A]$  from it. Since you have to additionally assume facts about the meaning of 'triangle' or 'square' to deduce a contradiction from a statement about the existence of a four-sided triangle or a round square, it will not be logically impossible in this narrow sense of 'logical possibility.' Usually, however, the phrase is used in a looser sense, where the sentences that are logically possible are those that do not entail contradictions on the basis of their logical form *plus* facts about the meanings the terms involved.
- 2 The tense logic of Arthur Prior discussed in Chapter 8 was based on this modal logic of C.I. Lewis.
- 3 Where the world in question may be the actual world.
- 4 Lewis is not a great fan of the word 'concrete.' He doesn't think the term has a sharp and well-defined meaning for the reasons we discussed in Chapter 2. He introduces concreteness merely to emphasize that these other worlds are as real as our own and exist in the same sense as our own.
- 5 To say that a number is irrational is to say that it is incapable of being expressed as a fraction of two integers, e.g.,  $1/4$ ,  $2/3$ , etc.

- a See Armstrong (1989b). Lewis and Armstrong disagree about whether there are genuine possibilities involving alien (i.e., otherworldly, merely possible) properties.
- 6 See Text Box 8.5.
- 7 Harry Truman was the president of the United States from 1945 to 1953.
- 8 See also Penelope Mackie's *How Things Might Have Been* for an alternative account of how we may be misled into thinking that objects and organisms have their origins essentially.

# 11 Causation

## Learning Points

- Presents influential accounts of causation from the history of philosophy
- Introduces regularity, counterfactual, probabilistic, primitivist, and physical theories of causation
- Evaluates arguments for and against these theories
- Distinguishes and evaluates the targets and goals of a metaphysical account of causation

## 11.1 Causation in the History of Philosophy

Most of us think that our world is full of causes and effects. Things don't just happen randomly, but often, if not most of the time, things are the way they are, events happen, because of other things or events that are their causes. A kitchen window is shattered into a hundred pieces and a rock is found lying outside. A cookie jar is empty and crumbs are found scattered beneath it. These are not random events but have explanations, causal explanations.

Causation is what the philosopher David Hume (1711-1776) called 'the cement of the universe.' It is what ties together the various disparate things that happen. Although much of contemporary metaphysics concerns questions of ontology, since the beginning of our discipline, philosophers have also been interested in questions about the connection between the various objects and events in our world and what are the reasons things are the way they are. If one seeks a complete account of the nature of our universe and what it is like, it doesn't seem sufficient to merely list the types of entities there are: the objects, properties, and events. We want to know as well the nature of the relations between these objects and events, which ones are tied causally to which others, and what these causal connections come to.

In Chapter 9 on persistence, we saw that Aristotle appealed to the distinction between matter and form to explain the behavior of material objects, how it is that they are able to survive through changes and in what these changes consist. For Aristotle, ordinary changes consisted in a substance, some matter, coming to take on a new form. But Aristotle too thought that more ingredients were required to provide complete explanations of why things are the way they are. A complete explanation should provide an account of an object's so-called efficient cause and also its teleological or final cause.

#### **Final (or teleological)**

**cause:** the purpose or goal for which an object exists or why it is the way it is at a given time

**Final or teleological causes** concern an object or event's *telos*, its purpose or goal. For example, the final cause of a window is to let in light. The final cause of my riding my bike today is to get to school. Today, especially due to the influence of naturalism and general skepticism about purposes in scientific explanations, many metaphysicians tend to deny that objects, especially natural objects, have final causes. Our best scientific theories tend not to appeal to purposes or goals to explain why things happen or why objects are the way they are, and so the naturalist argues, neither should we appeal to teleological causes in metaphysics. Philosophers of a more theistic persuasion tend to be less skeptical about final causes, however the present chapter will focus on the species of causation that is most commonly discussed by metaphysicians on both sides of the naturalist/theist divide. We will set aside final causation, although it is certainly an interesting issue whether objects do indeed possess purposes or final causes that can explain why they are the way they are.

**Efficient cause:** what brings an object or event into being

Contemporary discussions of causation typically concern something more like Aristotle's efficient causation. An **efficient cause** of an object or event explains what in its history brought that object or event into being. Usually, barring exotic cases like time travel, to find an object or event's efficient cause, one looks into the past. The efficient causes of objects are the processes that brought them into being: for artifacts, some kind of manufacturing process, for humans, the coming together of a sperm and an egg. Events' efficient causes are to be found in other events or actions. Two events involving one naughty child may be the efficient cause of both a window shattering and the emptying of a cookie jar.

#### **TEXT BOX 11.1**

#### **The Relata of Causal Relations**

Causation is typically thought to be a relation between entities, but entities of which kind? What types of entities are causes and what types of entities are effects? A common view is

that the relata of causal relations are events. For example, we say one event, the assassination of Franz Ferdinand, caused another event, World War I. Or we may point to a particular lightning strike (one event) as the cause of a particular wildfire (another event). There are different views about the nature of events, but the two most common are those associated with Donald Davidson (1970) and Jaegwon Kim (1976). According to Davidson, events are a species of concrete particular that involve changes taking place at particular spacetime regions, e.g., a window's breaking at a particular place and time. According to Kim, events are the exemplification of properties by objects at particular times, e.g., a window's having the property of breaking at a particular time.

Some have argued that a better way to think about causation is as a relation between tropes or property instances. Given the similarity between tropes and Kimean events, most don't take this view to be all that different from the standard view that the relata of the causal relation are events.

Others, such as D.H. Mellor (1938-2020) argued that it is facts, not events, that are causes (Mellor 1995). Whereas events are things that *happen*, facts are things that *are the case*. Facts, moreover, are often thought to be more finely individuated than events. What this means is that where there is only one event, one might think there are two or more corresponding facts. Take, for example, the event that consists of a certain ball going through a particular net at a time *t*. This one event might be described in several different ways and thus taken to correspond to several different facts:

It being the case that this ball went through the net at time *t*.

It being the case that a shot was scored at time *t*.

It being the case that the game-winning shot was scored at time *t*.

A more controversial position is that the relata of the causal relation are substances rather than events. For example, in certain cases, like that of free actions, some think it is important that particular kinds of substances be causes: human agents. This is the agent causation view. However, many remain skeptical about whether it truly makes sense to cite substances as causes (or effects). For consider a case of causation that is especially interesting to defenders of substance or agent causation: a case in which a politician makes a free decision to cast a certain vote, thus raising her arm. A defender of events as causes would likely say that the cause in this case was the politician's decision to vote, while a defender of substances as causes would say it was the politician herself that was the (or at least a) cause of her arm's rising. But if this is right, and it was the politician herself, not any particular decision she made or any other event occurring in her mind or body, then why was it that the arm raising occurred at that particular time and not earlier? For, she didn't come into existence suddenly at the time of the arm going up, but presumably existed for many years before that. If it is the substance herself that is the cause and not any particular decision or event on her part, then why doesn't the effect come into existence as soon as the substance exists?

Table 11.1 Distinguishing Objects and Events

Objects	Events
a particular window	a window's shattering at time $t$
a particular rock	a rock's striking a window at time $t-1$
Michelangelo's <i>David</i>	<i>David's</i> having a chip on his toe in 1991
COVID-19	The COVID-19 pandemic

## 11.2 Hume's Empiricism

In this section, we will consider a puzzle about the nature of causation from a later period in the history of philosophy that continues to inform discussions of the topic today. This puzzle concerns how we could ever learn about causal relations.

This philosophical problem was raised by David Hume. Hume was a defender of **empiricism**. This is the view that our knowledge and understanding of the world comes entirely from sensory experience. Although earlier philosophers, from Plato and Aristotle to Descartes and Leibniz, thought that some of our knowledge of the world is *a priori*, is available to us prior to experience by pure reasoning, empiricists like Hume were skeptical that we had any ideas or knowledge of the world that does not come from experience.

Hume held a particularly strict form of empiricism in which all of our ideas, the concepts we have that structure our thoughts and what we can know, are built up out of copies of simple sense impressions. For example, we can form the idea of a blue thing or a square thing, of a loud thing or a soft thing, because we have sense impressions of these kinds. The simple ideas (of blueness or squareness, loudness or softness) are copies our minds make of these basic sense impressions. For Hume, it is not that we cannot form ideas of things we haven't actually seen with our eyes, or heard or smelled or tasted or felt. We can, but any new ideas we form must ultimately be complex ideas built up out of the ideas of things we have already experienced. For example, we can easily come to have the idea of a unicorn, even though no one has ever seen one, by combining the ideas of things we have seen: a horse and a horn. One can easily have the idea of a one-mile-wide golden sphere as well, even if one hasn't ever seen one of those, by combining the simple ideas of goldness and roundness and being one mile wide.

If this form of empiricism is appealing to you as it was to Hume, then you might start to worry, as he did, about how one could ever form the idea of a causal relation between objects. Surely we can perceive objects and events and if we want to talk in Hume's way, we can say we form sense impressions of them, but do we ever directly perceive the causal relations

**Empiricism:** the view that our knowledge about the world comes entirely from experience

between objects and events? Is there a sense impression that we form of a causation itself? To take a silly example, say your friend sits across a table from you and quickly drinks a pint of beer. He then becomes red in the face and stumbles around drunk. This is a clear case of causation if anything is: his drinking causes his redness and stumbling. But although you may perceive his drinking and perceive as well his subsequent redness and stumbling, do you perceive the causal link between the drinking and the redness and stumbling? What would that even look like? But then if Hume is right and all of our ideas are built up out of simpler ideas copied from sense impressions we've had, then how is it possible for us to ever have the idea of the causal relation linking the drinking and the redness and stumbling?

A key part of the trouble the concept of causation makes for the empiricist is that causation is typically thought of as a necessary connection between events. Effects seem to be events that follow as a matter of necessity from their causes. An effect doesn't just follow a cause, it *has to follow*, once the cause occurs. When your friend quickly drinks a pint of beer, then if this drinking is a cause of his subsequent redness and stumbling, it is a sure thing, the universe guarantees, that he will get red and stumble around once he drinks the beer.<sup>1</sup> But although we may perceive the effect to follow the cause, it is hard to see how we may perceive it following the cause *with necessity*, as the cause really making the effect happen rather than simply happening before it. As Hume puts it in his *Enquiry Concerning Human Understanding*:

When we look about us towards external objects, and consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connexion; any quality, which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find, that the one does actually, in fact, follow the other. (VII.6)

There is a lively historical debate about what exactly Hume thought here and whether he thought that causes indeed must be events that necessitate their effects. Those philosophers who think that Hume held on to this position have called the view about causation they think he held 'skeptical realism.' This is the view that although causal relations really do exist in our world and they are necessary connections between events, we can't really understand this notion of causation and so the nature of causal relations will always be out of our cognitive reach. We can only predict with likelihood which events will follow which; we can't see the hidden or 'secret' connections between what happens.<sup>2</sup>



A far more common position one finds among historians of philosophy, however, is that Hume ended up rejecting the idea that causation is a necessary connection between events. Instead, they argue, Hume found a way in which our concept of causation may arise out of a combination of simpler ideas derived from sense impressions. Since there is no problem with our observing how one event may follow another, one might use such complex ideas of one event following another to build up to a concept of causation. A natural thought is that one may view our idea that one type of event is the cause of another type of event as constituted by our idea of the *regular succession* of events of these kinds. If an event of a given kind follows another in just one case, we might not view the sequence as indicating a causal link, but if this happens regularly (that is, time after time, again and again), then this would seem to suggest that what we are seeing is a case of causation. Returning to the previous example, we may have observed the quick drinking of a pint of beer to be followed by redness and stumbling not just in this one circumstance, but also in several instances in the past. We thus observe that redness and stumbling follows drinking with regularity. If so, we may infer from this that quickly drinking pints of beer is a cause of redness and stumbling.

This suggests an empiricist analysis of the concept of causation. An event of type A (say a drinking) is a cause of an event of type B (say a stumbling) just in case events that are As are regularly followed by events that are Bs. This analysis of the concept of causation involves replacing the idea of causation as a necessary connection with the more empiricist-friendly idea of causation as the regular succession of events. This is a simple version of the **regularity theory of causation**. Hume is often thought to be the first regularity theorist.<sup>3</sup> To say that a particular event *a* is the cause of another particular event *b*, on this view, is to say that these events *a* and *b* both occur, *a* is an A-type of event, *b* is a B-type of event, and A-type events are regularly followed by B-type events.

If this is on the right track, one might wonder why or how it is that anyone ever thought that causes necessitated their effects in the first place. How did we ever get confused into thinking we have the idea of necessary connections between events? Where did this idea come from? To answer this question, Hume proposed that after we see a regular sequence of events – an event of type B again and again following an event of type A, our mind comes to be led to have the expectation, when first observing an A-type event, that a B-type event is about to occur. This impression we have, this feeling upon experiencing one event that another is about to occur, is where we get the idea of something like a necessary connection between events. We mistakenly think this impression in our minds corresponds to some external necessitating force between the events themselves.

**Regularity theory of causation:** a theory of causation that explains causal relations in terms of the regular occurrence of patterns of events

**EXERCISE 11.1****Copying Ideas from Experiences**

Do you agree with Hume that we do not directly observe causal relations between events? What sort of examples might suggest otherwise? Besides causation, what are some other cases of concepts that we seem to have that do not appear to be built up out of ideas copied from sense impressions?

It is rare to find a philosopher today subscribing to the strict form of empiricism we find in Hume, that all of our ideas are built up out of simpler ideas copied from our sense impressions. There are certainly still questions about whether we should take causation to be a necessary connection between events, and if not, what other conceptions of causation are possible. But the worries about causation as a necessary connection don't arise for exactly the same reasons as they did for Hume. For example, a naturalist might think that we are able to grasp many new kinds of concepts through mathematical and scientific theorizing, and that this generally doesn't involve building concepts out of simpler ideas copied from sense impressions. Yet, the naturalist might still worry that concepts like necessity don't appear in the natural sciences, and so perhaps should be avoided in our best philosophical theories. Other philosophers might be fine with using the concept of a necessary connection between events, but wish for a reductive analysis that can explain these concepts in terms of more fundamental notions that are better understood. We turn now to consider the prospects for such a reductive analysis.

**11.3 Three Reductive Theories of Causation**

Reductive theories of causation are theories (like Hume's) that provide accounts of what it is for one event to cause another, where these accounts are formulated entirely in noncausal terms.<sup>4</sup> In this section, we will discuss the three most common kinds of reductive theory of causation: regularity theories of causation, counterfactual theories of causation, and probabilistic theories of causation.

There are many forms a regularity theory of causation may take, starting with the simple version discussed by Hume:

*Simple regularity theory:* an event *a* of type *A* causes an event *b* of type *B* just in case *a* and *b* actually occur and *A*-type events are regularly followed by *B*-type events.

We can see this is a *reductive* analysis of causation since the explanation, what is on the right side of the 'just in case,' makes no use of causal notions. The simple regularity theory is a good first try at a regularity theory of causation, but one might worry that mere regular succession isn't enough to make for genuine causation. For example, it is undoubtedly the case that every time your favorite football team has ever won a game, the sun has risen the following day. Your team's winning is regularly followed by the sun's rising. And yet it doesn't follow that the cause of the sun's rising on those days is your team's win.

A natural thing to say in response to a case like this is that causation can't be mere regular succession, but must involve something more. The effect must somehow follow the cause not just as a matter of coincidence but rather as a matter of the cause's *entailing* the effect. But what sense of entailment do we want? Logical entailment seems too much to require. (Indeed this would seem to bring us back to the idea of causation as a necessary connection between events.) To see why logical entailment between cause and effect doesn't anyway seem required for causation, consider the following paradigm cases of causation:

a lightning strike at time *t* causes a fire at time *t'*

a rock thrown at a window at time *t* causes that window to break at time *t'*

In neither case is there a logical entailment between the cause and the effect. If there is a relationship between lightning strikes and fires, or rock-throwings and window-breakings, it is not a logical relationship, but a relation of some other kind.

Instead, many regularity theorists have suggested that causation requires that the effect be entailed by the cause not on its own as a matter of logic, but as a result of the laws of nature. For example, a regularity theory of this kind might look like:

*Nomic regularity theory*: an event *a* of type *A* causes an event *b* of type *B* just in case *a* and *b* actually occur and the laws of nature imply that *A*-type events are regularly followed by *B*-type events.

The laws of nature don't entail that the sun will rise tomorrow, nor do they imply that if your football team wins today, the sun will rise tomorrow. So this fix works to rule out the sort of case that is a problem for the simple regularity theory. And, in general, it works to avoid counting cases of succession by coincidence as cases of causation. The nomic regularity theory also seems to nicely capture those cases we ordinarily think of as cases of causation - it does seem to follow as a matter of natural law that drinking leads to redness and stumbling, that dropping objects causes them to fall, that lightning strikes lead to fires, that throwing rocks at windows causes those windows to break, and so on. Many philosophers

have endorsed a version of the regularity theory that builds in entailment according to the laws.<sup>5</sup>

Yet in the 1970s, David Lewis pointed out several difficult issues for regularity theories of causation that caused many to reconsider the position. Instead, Lewis proposed a distinct kind of reductive theory of causation, a **counterfactual theory of causation**.<sup>6</sup> Lewis's theory has been extremely influential ever since, not just in philosophy, but in how causation is thought about in psychology and the social sciences as well.<sup>7</sup>

Two of the problems Lewis raised for the regularity theory were the problem of epiphenomena and the problem of preemption.<sup>8</sup> **Epiphenomena** are events that are the results of other events, but that have no effects of their own. What Lewis argued was that it could be the case that an event *a* may satisfy the requirements of a regularity theory of causation to count as a cause of another event *b*, while *a* actually isn't a cause of *b*. This could happen if *a* were an epiphenomenon traceable to some other event in *b*'s causal history. For example, let's consider again the case of your friend's drinking and this leading to his turning red and stumbling. The causal structure of this situation may be represented as shown in Figure 11.1, with the dots representing events and arrows representing relations of causal influence.

In this case, the drinking (*c*) is a cause of both the reddening (*a*) and the stumbling (*b*). After the drinking, the reddening occurs first and then the stumbling. But the stumbling is not caused by the reddening. Despite this fact, a regularity theory would take the reddening to be a cause of the stumbling. Because the laws of nature entail the drinking will be followed by first a reddening and then a stumbling, the laws entail a reddening will be followed by a stumbling. And this is what the regularity theorist says is sufficient for the reddening to be a cause of the stumbling. But the reddening is an epiphenomenon. It is caused by something earlier in the causal chain (the drinking) but it doesn't have any effects of its own. So, the regularity theory has a problem. It counts epiphenomena as causes.

**Counterfactual theory of causation:** a theory that reduces facts about causation to facts about what would have happened in various counterfactual circumstances

**Epiphenomenon:** an entity that lacks causal efficacy

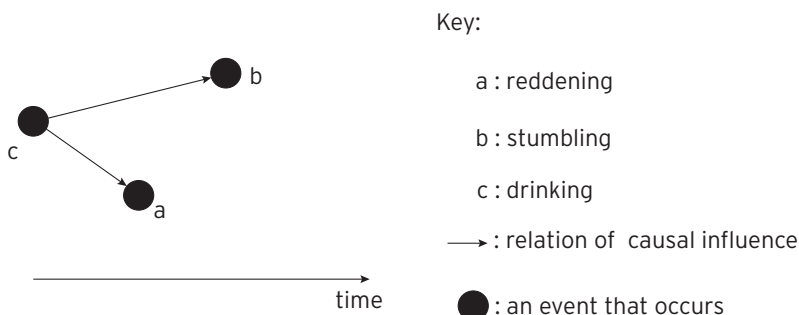


Figure 11.1 The Problem of Epiphenomena

Another problem for regularity theories is the problem of preemption. Suppose two naughty children, Billy and Suzy, are trying to break a window by throwing rocks at it. Billy throws his rock first, aiming carefully and throwing with enough force to break the window. But just as Billy releases the rock from his hand, Suzy throws hers in just such a way as to knock Billy's off its trajectory. Suzy's rock bounces off of Billy's into the window, thus breaking it. In this case, the obvious thing to say is that it was Suzy's throw that was a cause of the window's breaking. Billy's throw could have been a cause, but in fact it wasn't, since it was *preempted* from being a cause of the window's breaking by Suzy's throw. This is represented in Figure 11.2. Steps on the causal chain that would have led from Billy's throw to the window's breaking have been inhibited by Suzy's throw.<sup>9</sup>

Let's first see why preemption presents a problem for the regularity theory. What makes this a case of preemption is that the preempted cause, Billy's throw, on its own seems to have what it takes to cause the window's breaking. It occurs, the effect occurs, and events just like Billy's throw are regularly followed, according to the laws, by window-breakings. And so according to the regularity theory, it counts as a cause of the window-breaking. But it is not a cause of the window-breaking. Only Suzy's throw is. Thus, the regularity theory has another problem. It counts preempted causes as causes.

The account Lewis proposes as a replacement to the regularity theory of causation begins with the notion of counterfactual dependence between events. Recall from Chapter 10, a counterfactual is a conditional in which one supposes that something occurred differently than it actually did. Some examples of counterfactuals are:

If kangaroos didn't have tails, they would topple over.

If Paris were in England, then people would speak English there.

If Suzy hadn't thrown her rock, then the window would still have broken.

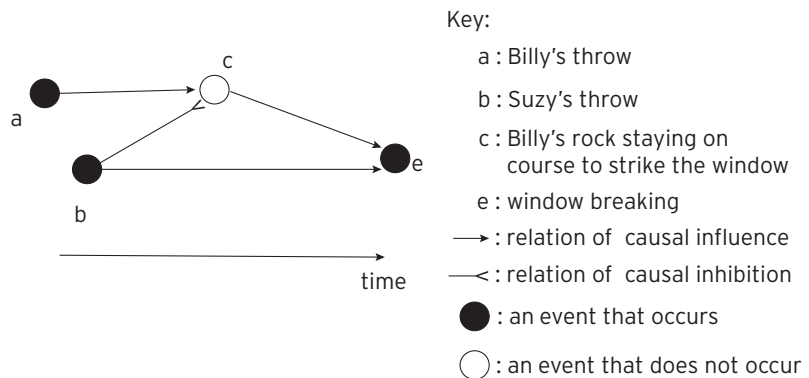


Figure 11.2 The Problem of Preemption

Lewis proposes to analyze causation in terms of counterfactual dependence, where one event *e* counterfactually depends on another event *c* just in case had *c* not occurred, then *e* would not have occurred either. Note: Lewis analyzes causation *in terms* of counterfactual dependence. He does not analyze causation *as* counterfactual dependence. Lewis's original account of causation from his 1973 paper "Causation" is the following:

*Lewis's 1973b counterfactual theory of causation:* *c* causes *e* just in case the events *c* and *e* both actually occur, and there is a chain of counterfactual dependence running from *c* to *e*.

What it is for there to be a chain of counterfactual dependence running from *c* to *e* is for there to be some sequence of events starting with *c* and terminating with *e* such that every event in the sequence counterfactually depends on the event that immediately precedes it in the sequence. In some cases, this chain will only involve *c* and *e*. In this case, *c* will be a cause of *e* in virtue of both *c* and *e* occurring and the following counterfactual being true:

If *c* hadn't occurred, then *e* wouldn't have occurred either.

But in other cases, there may be no direct dependence of *e* on *c*.

Actually, cases of causal preemption are good cases to use to illustrate why counterfactual theorists like Lewis think causation should be analyzed in terms of a chain of counterfactual dependence. We have already noted that the case of Billy and Suzy is a problem for the regularity theory of causation since there is regular nomic succession of rock-tosses like Billy's and window-breakings and yet Billy's throw is not a cause of the window's breaking.

But cases of preemption also show why a simple counterfactual theory of causation couldn't be true either.

*Simple counterfactual theory of causation:* *c* causes *e* just in case the events *c* and *e* both actually occur, and *e* counterfactually depends on *c*.

The simple counterfactual theory would get it right that Billy's throw is not a cause of the window breaking. The following counterfactual after all is false:

If Billy hadn't thrown his rock, the window wouldn't have broken.

This is false because if Billy hadn't thrown his rock, Suzy still would have been there to throw hers. And so even if Billy hadn't thrown his rock, the window would have broken. This is a good result for the simple counterfactual theory. But there is also a bad result for this theory, namely, it

fails to count Suzy's throw as a cause of the window breaking. For this counterfactual is also false:

If Suzy hadn't thrown her rock, the window wouldn't have broken.

This counterfactual is false because of the presence of Billy and his throw. If Suzy hadn't thrown her rock, the window still would have broken because of Billy's throw.

This is one of the main reasons why Lewis complicates his counterfactual theory of causation to make it require not counterfactual dependence between *c* and *e*, but instead a chain of counterfactual dependence between *c* and *e*. Although there is not direct counterfactual dependence of the window's breaking on Suzy's throw, there is a chain of counterfactual dependence leading from Suzy's throw to the window breaking. We can see this chain by adding more detail to the previous figure. See Figure 11.3.

Even though the following counterfactual is false:

If Suzy hadn't thrown her rock, the window wouldn't have broken,

the following counterfactuals are both true:

If Suzy hadn't thrown her rock, then her rock wouldn't have stayed on course to strike the window (i.e. if *b* hadn't occurred then *d* wouldn't have occurred).

If Suzy's rock hadn't stayed on course to strike the window, the window wouldn't have broken (i.e. if *d* hadn't occurred, then *e* wouldn't have occurred).

Since these counterfactuals are both true, there is a chain of counterfactual dependence linking *b* with *e*. And so Lewis's account counts Suzy's throw as a cause of the window's breaking.

As has been mentioned, many both inside and outside philosophy have been attracted to the idea of a counterfactual theory of causation. The idea is so natural that even Hume (although he seems to have ultimately endorsed a regularity theory) says that causation is the idea

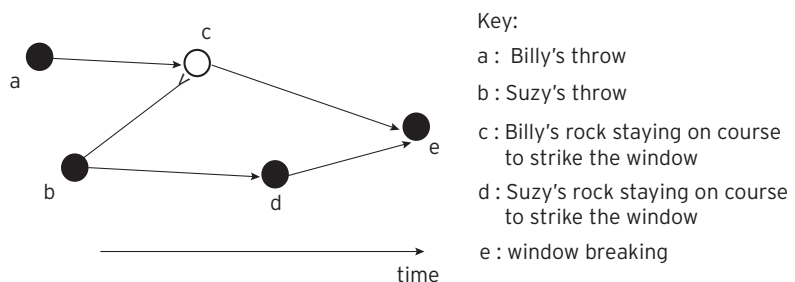


Figure 11.3 Billy and Suzy (Detail)

that if *c* hadn't occurred, *e* wouldn't have occurred (*Enquiry* VII.29). Still, counterfactual theories are not without their problems, and especially in the 1990s and 2000s there was a cottage industry devoted to producing counterexamples to Lewis's theory. The philosophers responsible for these counterexamples often offered more sophisticated counterfactual theories of causation tailored to avoid the counterexamples. Indeed, the definitive version of Lewis's "Causation," included in his 1986 *Philosophical Papers Volume II*, includes a series of postscripts in which he discusses many of these problem cases and proposes a more sophisticated model (one he also ultimately rejected) to repair the damage.<sup>10</sup>

## EXERCISE 11.2

### Regularity and Counterfactual Theories of Causation

Determine what the following four theories of causation would say about each case: (i) the simply regularity theory, (ii) the nomic regularity theory, (iii) the simple counterfactual theory, (iv) Lewis's counterfactual theory.

- A. Every time Tony wears a white shirt, he spills coffee down the front of it. This happens often when he wears shirts of other colors as well, but it happens every time he wears a white shirt. Is Tony's wearing a white shirt a cause of his spilling coffee on himself?
- B. White and Scarlet, two assassins, have both been hired to murder Mr. Body. Both are skilled and reliable at carrying out their missions. On this occasion, White finds their target first. As Body is just about to die of strangulation from White's rope, Scarlet enters the room and shoots Body in the chest, stopping his heart. Was Scarlet's firing her gun a cause of Mr. Body's death?
- C. Michael is set to compete in a major pickleball tournament. The night before the finals, he trips and falls, breaking his right hand. As a result he is only able to use his left hand. This is disappointing, since Michael is right-handed and not used to using his left hand to play pickleball. Miraculously, playing with his left hand only, Michael wins the championship. Is Michael's tripping and falling the night before the finals a cause of his winning the championship?

Alongside regularity and counterfactual theories, there is a third kind of reductive theory of causation that has been influential. Some prefer to analyze causation not in terms of regularities, laws, or counterfactuals, but rather in terms of probabilities. One simple version of a probabilistic theory of causation says that causes are probability-raisers. More precisely:

*Probability-raising theory of causation:* *c* is a cause of *e* just in case *c* and *e* both actually occur, and the occurrence of *c* raises the probability of *e*'s occurring.



This is an intuitive and informal formulation of the idea. Usually, however, the theory is formulated using the formal notion of conditional probability, where the probability of *e*'s occurring conditional on *c*'s occurring is written  $P(e|c)$ . Once we have this notation, we can say that the occurrence of *c* raises the probability of *e*'s occurring just in case  $P(e|c) > P(e)$ . In other words, *c* raises the probability of *e* just in case the conditional probability of *e* given *c* is higher than the antecedent probability of *e*'s occurring.

The probabilistic view has appealed to philosophers for several reasons. First, one might think it is a way of capturing the intuition behind regularity accounts without requiring something so strict as entailment between a cause and an effect. To return to the earlier case, we might think the drinking caused the stumbling even if the laws didn't strictly imply that the stumbling would happen, it only made it much more likely. The probabilistic view is also popular due to the fact that probabilistic or statistical correlation seems to be the concept of causation that is operative in many of the sciences (both the natural and social sciences). In medicine, something is taken to be a cause of a disease if it raises the probability of one's developing the disease. In sociology, something is taken to be a cause if it is a statistically significant factor.

Still, probabilistic theories are not without their difficulties. First, it is not always clear how to apply the concept of probability to events, even events that we very clearly believe have causes. For example, is it really possible to assign any definite probability to your friend's stumbling around given his drinking or not drinking a pint of beer? Even if we could find a way to agree on a definite probability to assign your friend's stumbling around on a particular occasion, this probability assignment would likely depend on a lot of other assumptions we make about the case and thus how we are thinking about it. This is what is known as the **reference class problem**. The probabilities we assign to the occurrence of particular events do not seem to be objective features they possess, but instead seem to be determined by our contingent ways of conceptualizing those events on a given occasion. But causation is supposed to be an objective phenomenon concerning the relations between events. Whether one event is the cause of another doesn't seem to depend on how we think about these events, or whether we think about them at all.

#### Reference class

**problem:** the problem that the probability we assign to an event's occurring seems to depend on our way of conceptualizing it (placing it against a reference class) on a given occasion; since the reference class we use can depend on the context, this makes it difficult to say what is the probability of an event's occurring

#### TEXT BOX 11.2

##### Van Fraassen's Cube Factory

Bas van Fraassen once used a clever example to illustrate the fact that which probabilities it is reasonable to assign to a type of event's occurring depends on how we choose to think about the situation (1989, p. 303). Suppose there is a factory designed to produce cubes of various

sizes. We want to know the probability on a particular Monday that the factory will make a cube with a particular size. Here is a description of the situation:

This factory is equipped with machinery to make cubes with edges of any length up to 2 cm. For any length up to 2 cm, there is an equal chance on Monday that the factory will manufacture cubes with edges of that length. *What is the probability that on Monday the factory will manufacture cubes with edges less than 1 cm in length?*

Think about this. Since the factory can make cubes with edges of any size up to 2 cm, and half of the possible lengths up to 2 cm are  $<1$  cm, it would seem the correct answer to this question is  $\frac{1}{2}$ . The probability that on Monday the factory will manufacture cubes with edges  $<1$  cm in length would seem to be  $\frac{1}{2}$ .

But here are two other situations we might describe and questions we might ask:

This factory is equipped with machinery to make cubes with sides of any area up to  $4\text{ cm}^2$ . For any area up to  $4\text{ cm}^2$ , there is an equal chance on Monday that the factory will manufacture cubes with sides of that area. *What is the probability that on Monday the factory will manufacture cubes with sides less than  $1\text{ cm}^2$  in area?*

This factory is equipped with machinery to make cubes with any volume up to  $8\text{ cm}^3$ . For any volume up to  $8\text{ cm}^3$ , there is an equal chance on Monday that the factory will manufacture cubes with that volume. *What is the probability that on Monday the factory will manufacture cubes less than  $1\text{ cm}^3$  in volume?*

But now you should be scratching your head. The answers to the second and third questions would appear to be different from the answer to the first. The answer to the second question seems to be  $\frac{1}{4}$ . The answer to the third question seems to be  $\frac{1}{8}$ . But the trouble is that each paragraph is describing the same scenario and asking the same question about that scenario. To say that our factory can make cubes with edges up to 2 cm is to say the same thing as to say our factory can make cubes with sides up to  $4\text{ cm}^2$  in area, and this is to say the same thing as to say our factory can make cubes with volumes up to  $8\text{ cm}^3$ . And to ask for the probability of the factory producing a cube with edge length 1 cm or less is to ask for the probability of the factory producing a cube with sides of area  $1\text{ cm}^2$  or less, which is again just another way to ask for the probability of the factory producing a cube with volume  $1\text{ cm}^3$  or less.

Which probability one assigns to a given situation then would seem to depend on the way one describes that situation and the concepts one is applying to consider the case. This is the reference class problem. It is the problem of finding the correct reference class against which to compare the given possibility – in this case, one with cubes with other possible edge lengths, or cubes with other possible side areas, or cubes with other possible volumes.

Beyond the reference class problem, philosophers have also noticed that probabilistic theories of causation seem threatened by the same problems of epiphenomena and preemption that plagued regularity theories. Returning to the problem of epiphenomena, recall that in the drunk friend case, the redness is regularly followed by stumbling, but at the same

time is clearly not a cause of the stumbling. The redness is an epiphenomenon. And here we may note that the friend's redness may also raise the probability of his stumbling, but that in no way makes it follow that it is a cause of the stumbling. Instead, both the redness and the stumbling are both effects of a common cause: the drinking. The probability-raising theory also seems to get the wrong result in the Billy and Suzy preemption case. Billy's throw both occurs and raises the probability of the window's breaking. Yet, despite this fact, it is not a cause of the window's breaking. It is Suzy's throw instead that causes the breaking. The probability-raising theory thus seems to get the wrong result.

A final problem we will consider for probabilistic theories of causation is that there appear to be cases in which *c* causes *e* but the occurrence of *c* seems definitely to lower, not raise, the probability of *e*'s occurring. Lewis considers a case in which we have two systems in place that are capable of triggering a given event *e*. System A is reliable and will raise the probability of *e*'s occurring to 99 percent. System B is much less reliable and makes the probability of *e*'s occurring 50 percent. Suppose one throws a switch to turn off system A and turn on system B. Later, system B works and, as a result, the event *e* happens. The throwing of the switch lowered the probability of *e*'s occurrence, but it is still intuitively a cause of *e*'s occurrence. Cases like this suggest that sometimes a cause is something that doesn't raise the probability of an event's occurring. Whether this means the probabilistic theory should be thrown out altogether, revised, or supplemented with additional criteria is an open question in the philosophy of causation.

## 11.4 An Objection to Reductive Theories of Causation

We have now seen three approaches to developing a reductive theory of causation, that is, one that tries to produce an account of causation in non-causal terms: (a) regularity theories that try to analyze facts about causation in terms of facts about regular sequences of events, and perhaps laws, (b) counterfactual theories that try to analyze facts about causation in terms of facts about what would have happened under circumstances in which the putative cause did not occur, and (c) probabilistic theories that analyze causation in terms of probabilistic notions. These accounts have been influential and are all still very prominent, however some philosophers remain skeptical of the very idea of reducing facts about causation to noncausal facts and it is worth seeing at least one example to get the flavor of these objections.

One set of objections comes from the philosopher Michael Tooley (for example, in his 1990 paper "Causation: Reductionism vs. Realism"). Tooley defends a **primitivist theory of causation** – a theory according to which causal facts are not reducible to any noncausal facts, including facts about

**Primitivist theory of causation:** a theory according to which causal facts are not reducible to any noncausal facts, including facts about regularities, laws, counterfactuals, or probabilities

regularities, laws, counterfactuals, or probabilities. Tooley himself calls his view 'realism about causation,' but this is of course contentious. Davidson, Lewis, and the others who have developed and endorse reductive theories of causation consider themselves "realists" as well about causation, even if they think causal notions are ultimately explainable in more basic terms. 'Primitivism' is thus a more neutral term for Tooley's position.

One of Tooley's cases involves a simple world with only two fundamental laws.<sup>11</sup> The first law of this world is:

(L<sub>1</sub>) For any object x, x's having property P at time t causes x to acquire one of property Q or property R at time t'.

The second law is:

(L<sub>2</sub>) For any object x, x's having property S at time t causes x to acquire one of property Q or property R at time t'.

Note that these are both indeterministic laws. They don't say with necessity which type of events follow from which others, but make a weaker claim about what could follow (either Q or R).

Suppose a single object has property P at one time and then later goes on to acquire Q (or R). We may then infer using any of our reductive theories and L<sub>1</sub> that it was this object's having P that caused it to acquire Q (or R). And similarly, if at one time an object has property S and then later acquires Q (or R), we may infer, using any of our reductive theories and L<sub>2</sub>, that it was this object's having S that caused it to acquire Q (or R). The trouble, Tooley points out, arises when there is a situation like the following:

<i>Time t</i>	<i>Time t'</i>
Object a has properties P and S.	Object a has properties Q and R.

In this situation, some real questions arise. The laws entail that the instantiation of P causes the instantiation of Q or R. The laws also entail that the instantiation of S causes the instantiation of Q or R. But in this situation, was it the instantiation of P by a that caused a to possess Q? Or was it the instantiation of S by a that caused a to possess Q? And similarly, we can ask about a's instantiation of R at t'. Was it a's possession of P or of S that caused a to have R at t'?

Tooley's point is that the facts about the laws underdetermine the facts of causation in this case. And so the facts about causation cannot be analyzed in terms of nomic regularities. Nor does he think they may be adequately analyzed in terms of counterfactuals. Laws L<sub>1</sub> and L<sub>2</sub> may also supply us with evidence about which counterfactuals are true. This still won't determine which it is, Pa or Sa, that is responsible for Qa, and which

of Pa or Sa it is that is responsible for Ra. To argue against probabilistic theories, Tooley modifies his case to make the laws include facts about probabilities. E.g.,  $L_1$  might be modified to say that an object's having property P raises the probability of its later having Q or R.  $L_2$  might be modified to say that an object's having property S raises the probability of its later having Q or R. This, Tooley argues, would show that the facts about probabilities underdetermine the facts about causation. Thus, Tooley concludes, we ought to see the facts about causation as additional facts irreducible to any facts about regularities or counterfactuals or probabilities. The causal facts are not reducible to any more fundamental facts. Causal facts are "primitive."

### EXERCISE 11.3

#### Tooley's Objection to Reductive Theories of Causation

Select what you take to be the most promising reductive theory of causation. How should a defender of this theory respond to Tooley's alleged counterexample?

## 11.5 Process Theories of Causation

So far, we have introduced several different kinds of theories of causation: regularity theories, counterfactual theories, probabilistic theories, and primitivist theories. This leaves out one important class of theories of causation: process theories of causation. Process theories of causation try to give an account of causation in terms of the occurrence of physical processes.

In not all cases are defenders of process theories trying to analyze all cases of causation. Often, process theorists only claim to analyze the kinds of processes involved in physical causation. However, some recent process theorists of causation do attempt to do more than describe the sort of processes involved in physical causation. They go so far as to try to explicate what causation is in all cases in terms of physical processes. For this reason, their theories are often called 'physical theories of causation.'

Probably the physical theory of causation that is most discussed today is the one that was proposed in a 2000 book by Phil Dowe called *Physical Causation*. The account there was heavily influenced by previous work of Wesley Salmon (1925-2001) and Hans Reichenbach (1891-1953). Dowe's theory starts with an account of two central causal notions, that of a causal process and that of a causal interaction.

First, Dowe borrows the notion of a **worldline** of an object from special relativity. This is the path of any object through spacetime. It is in terms of this notion of a worldline and the additional notion of a conserved quantity from physics that Dowe develops a general theory of causation:

**Worldline:** the path of any object through spacetime

*A causal process* is a worldline of an object which manifests a conserved quantity,

*A causal interaction* is an intersection of worldlines which involves exchange of a conserved quantity,

*An exchange* is a case in which at least one incoming and at least one outgoing process manifest a change in the value of the conserved quantity.<sup>12</sup>

Conserved quantities are physical quantities that do not change their total values overtime. For example, the principle of conservation of energy says that the total energy of a closed system (one that does not interact with an outside environment) never changes over time. Over time, individual objects may gain or lose energy. But if energy is a conserved quantity, any loss in energy by one object must be made up for in the gain in energy by another object.

Dowe's account is widely taken to be the most sophisticated and well-worked out physical theory of causation available today. Still, it has met with many objections from different parties.

On the one hand, there are those philosophers who are skeptical in general of the project of understanding causation in terms of physical processes. Several metaphysicians have complained that physical theories ignore the many cases of causation involving not a physical process, but the *absence* of a physical process. For example, many believe there are cases of causation by omission, such as:

(1) Sally's failure to water the plants caused the plants to die.

There also appear to be compelling cases of causation of absence:

(2) Sally's wearing her seatbelt caused her not to be ejected from her seat.

There even seem to be cases in which both the cause and the effect involve the absence of a physical process:

(3) Sally's failure to send an invitation caused Luis not to attend the party.

In all of these cases, the causal claims seem to describe not a physical process involving the exchange of a conserved quantity, but the absence of a physical process. In these cases, since the process fails to occur,

there can be no exchange of conserved quantities as the process theory seems to require.

Usually, such cases are presented in support of a counterfactual theory of causation. In all of the above cases, we have corresponding counterfactuals whose truth does not appear to require the occurrence of any physical process:

(1CF) If Sally had watered the plants, the plants wouldn't have died.

(2CF) If Sally hadn't worn her seatbelt, she would have been ejected from her seat.

(3CF) If Sally had sent Luis an invitation, Luis would have attended the party.

The counterfactual theorist will argue that the causal claims (1) - (3) are true in virtue of the truth of the counterfactuals (1CF) - (3CF). Since there are genuine cases of causation that physical theories cannot explain, the physical theory must be inadequate as a theory of causation.

#### **EXERCISE 11.4**

#### **Causation Involving Absences and Process Theories of Causation**

Defenders of physical process theories typically have two ways of responding to objections appealing to cases of causation involving absences. They have the option of (a) arguing that the case in question is not a genuine case of causation, or (b) arguing that in the relevant case, there is a physical process one can point to that grounds the causal fact. For each of (1)-(3) above, decide how it is most appropriate for the physical process theorist to respond.

Dowe's theory has also faced objections from those who are in principle sympathetic to the project of developing a physical process theory of causation. Some philosophers of science have worried that Dowe's account fails to capture the full range of causal processes and interactions that one finds in our best physical theories. For example, modern physics describes some causal interactions that don't involve the crossing of worldlines. Objects typically interact with each other at a distance through the intermediary of fields without their worldlines actually intersecting. This suggests that more work is required to better understand physical causal processes. This is not a surprise. As physics continues to evolve, our physical theories of causation should evolve as well.

## 11.6 Two Projects in the Philosophy of Causation

One issue worth considering at this stage is the status of all of these competing accounts of causation with respect to one another. Often regularity, counterfactual, probabilistic, primitivist, and process theories of causation are presented as if they are competitors to one another, but there are many distinct aims that a theory of causation may be taken to have. With this in mind, it is important to keep in mind when evaluating a theory of causation what exactly that theory is trying to explain. It is possible that some of these theories may actually be compatible with one another, if they are trying to give accounts of different phenomena.

For example, it is very clear in Lewis's work on causation that what he is aiming at is an account that captures the truth of the majority of causal claims we assert in ordinary circumstances.<sup>13</sup> His counterfactual theory of causation is thus intended as an **analytic account** of causation, an account of what we mean when we say something of the form [c causes e]. Lewis acknowledges the possibility that his account might fail in some cases to track what we might ordinarily say or that there may be some cases (tricky, complicated philosophers' examples) where our intuitions just won't settle what the correct answer will be. But that is okay. What he is aiming for is an account that gets most of what we ordinarily say right and the weird, complicated cases can be left as a matter of "spoils to the victor." Whichever account handles the majority of cases, the cases where our intuitions are clear, gets to decide what it is right to say about the cases where our intuitions fail to deliver an obvious result.

It is widely agreed that counterfactual theories are well placed to address what it is we try to express when we assert causal claims. Most philosophers of causation today are sympathetic to what Carolina Sartorio (2005) has called "the difference-making intuition," the idea that causes are those events that make a difference to their effects. If e is counterfactually dependent on c, then this is to say that if c hadn't happened, then e wouldn't have happened either. This is one obvious way in which c could make a difference to e. If our ordinary talk about causation assumes causes are difference-makers, then this would also seem to give support to probabilistic theories of causation. For another way for c to make a difference to e is for the occurrence of c to raise the probability of e's occurring.

On the other hand, in his book, Dowe is clear that he is not trying to provide an analytic account, an account that captures what we mean when we make everyday causal assertions, even an account according to which most of the causal claims we make are true. Rather, Dowe is trying

**Analytic account:** an account of what we mean when we make some claim



to develop an accurate empirical account of causation. What he is trying to do is spell out what causation is in our world, according to our best scientific theories. Our best scientific theories can often tell us things that conflict with our ordinary intuitions, things we might frequently assert in everyday circumstances whether we know it or not. And so, it is clear that the project Dowe has in mind, as he develops his account of causation, is distinct from the project Lewis was working on.

It is thus possible that multiple kinds of theories may have a role in the ultimate, complete philosophical account of causation. We might learn from our best scientific theories that there are certain distinctive kinds of physical processes that are characteristic of causation at our world. And yet a distinct account may turn out to be appropriate when we attempt to capture the truth conditions of most of our ordinary causal claims. Many have thought that counterfactual theories of causation are best up to this task. There have certainly been many sophisticated counterfactual theories developed to meet the challenge. But it is still possible that no reductive theory will ultimately suffice to capture what we mean by our causal claims, which is what Tooley's arguments were intended to show.

### **Suggestions for Further Reading**

The two main theories of events with which metaphysicians work today may be found in Jaegwon Kim's "Events as Property Exemplifications" and Donald Davidson's "Events as Particulars." Aristotle's four causes (material, formal, efficient, and final) are presented in his *Physics*, Book II, Chapter 3, and his *Metaphysics*, Book V, Chapter 2. For an interesting defense of teleological causation, see John Hawthorne and Daniel Nolan, "What Would Teleological Causation Be?" Hume's theory of causation can be found in his *Enquiry Concerning Human Understanding*, sections IV-VII. For a recent discussion of whether we directly experience causal relations, see Susanna Siegel's *The Contents of Visual Experience*.

An excellent anthology collecting many of the classic papers discussed here including those by Davidson, Lewis, and Tooley, among others, is *Causality*, edited by Ernest Sosa and Michael Tooley. The 2004 anthology *Causation and Counterfactuals*, edited by Jessica Collins, Ned Hall, and L.A. Paul, contains many important papers on counterfactual theories of causation, including critiques by David Armstrong and Tim Maudlin, and later versions of the counterfactual theory offered, for example, by David Lewis, L.A. Paul, and Stephen Yablo.

For physical and probabilistic theories of causation, one may consult the papers by Wesley Salmon in the Sosa and Tooley volume. Predecessors

for the type of process theory found in Dowe's *Physical Causation* may be found in Reichenbach, *The Direction of Time*, and Salmon, "Causality without Counterfactuals."

There are two other important theories of causation that we did not have the space to examine here. One is the INUS-condition account of J.L. Mackie developed in his 1980 book *The Cement of the Universe*. This is a more sophisticated version of the regularity theory. Causes are taken to be "INUS-conditions," conditions that are insufficient but necessary parts of a condition which is itself unnecessary but sufficient for the result. Another kind of account of causation that has become very popular in recent years, especially among philosophers of science, is the interventionist account popularized by James Woodward in his 2003 book, *Making Things Happen*. Interventionism is a sophisticated version of the counterfactual theory that provides an account of what causes what in terms of facts about what the results of various interventions on a system would be.

## Notes

- 1 You might worry that strictly speaking, just because your friend quickly drinks a pint, this doesn't absolutely guarantee that he will get red and stumble around. Your friend could have eaten an entire loaf of bread immediately before the pint in an effort to soak up the alcohol he was about to drink. He could also have fallen asleep immediately after he drank it and never had the chance to stumble around. In either case, the defender of causation as a necessary connection between events will argue that to find the complete cause of the redness and stumbling, we need to fill in a bit more about the background conditions of the case (e.g., his starting with a fairly empty stomach while remaining awake).
- 2 Galen Strawson defends the view that Hume was a skeptical realist about causation in his book *The Secret Connexion*.
- 3 Hume also adds a contiguity constraint to his regularity theory. Causes must be contiguous to their effects, meaning they must occur at a spatial location near or next to their effects.
- 4 To be precise, one should make a distinction between theories that try to analyze *type causation*, what it is for one type of event to cause another type of event (e.g., for smoking to cause cancer, marriage to cause happiness), and those theories that try to analyze *token causation*, what is required for a token event to be the cause of another token event (e.g., for Jim's smoking to cause him to develop cancer, the marriage of Tony and Maria to cause their happiness). In this chapter, we will focus on theories of token causation.
- 5 One influential version appears in Davidson (1967b).
- 6 This theory was first developed in his paper from 1973b, "Causation." The account was later modified in the postscripts added to the paper in its 1986 reprinting in Lewis's *Philosophical Papers: Volume II*. Lewis modified his theory again in 2000 in his paper "Causation as Influence."
- 7 For example, the cognitive scientist Judea Pearl's influential book on the topic, *Causality*, explicitly acknowledges the influence of Lewis's theory.
- 8 It is worth noting that although counterfactual theories of causation are definitely much more common today as a result of Lewis's work, regularity theories

have not completely disappeared. Strevens (2007) offers a nomic regularity theory responsive to Lewis's concerns.

- 9 For more on the use and interpretation of these diagrams, see Ned Hall and Laurie Paul's, *Causation: A User's Guide*.
- 10 The book *Causation and Counterfactuals*, edited by Collins, Hall, and Paul, collects much of this work.
- 11 The example has been very slightly modified to simplify the presentation.
- 12 Dowe also supplements these definitions with an account of causal connection which we will not get into here.
- 13 It is thus, returning to a distinction we introduced in Chapter 5, a project in descriptive, rather than revisionary, metaphysics.

# 12 Free Will

## Learning Points

- Introduces the various conceptions of free will at stake in the contemporary philosophical literature
- Considers the putative conflict between free will and determinism, and various ways of resolving the conflict
- Assesses the case for and against determinism
- Explores the connection between free will and moral responsibility
- Considers views that deny that humans possess free will

## 12.1 What is Free Will?

Until now we have focused on metaphysical issues that concern objects considered in a very general way. We have discussed fundamental issues in ontology, fundamentality, the nature of time, modality, and causation. However, there are a host of other issues in metaphysics that are not quite so general but rather concern some important features of specific kinds of entities. One metaphysical issue that has been incredibly interesting to philosophers for centuries is whether it is possible for people to have free will. Are the actions of ordinary human beings like ourselves ever truly up to us? Do we ever have any control over the types of things we do and the kinds of people we are?

Typically, these philosophical questions arise out of worries that there exists some threat to our having free will. Sometimes the threat comes in the form of a God or other deity who, it is suggested, might be controlling our natures and our decisions about what to do. Other times, and this is the problem that will concern us here, the threat comes from the impersonal and universal laws of nature that govern everything that happens. The metaphysical issue then concerns how to reconcile the sense that we

are free agents with the existence of these laws. Or if this is not possible, the job of the philosopher becomes to assess the important implications the laws have for the way we must see ourselves and the relationships we have to our environments and other human beings.

So, the main issue in this chapter is whether in any important sense any of our decisions or actions are up to us. This is what it means to say an agent's action or decision is free – it is up to that agent. It comes from her and not from someone or something else. The philosopher Robert Kane has argued that there are different species of freedom we may distinguish: what he calls **surface freedom**, on the one hand, and a *deeper* form of freedom, or **ultimate freedom**, on the other.<sup>1</sup> Most metaphysicians who discuss the topic of free will agree that this distinction exists and indeed it is freedom in the deeper sense that we should care about.

**Surface freedom:** being able to act in such a way that one's desires are satisfied

**Ultimate freedom:** having the ability to not only satisfy one's desires but being the ultimate source of these desires

To see the distinction, imagine one day you have nothing special to do, no outside responsibilities demanding your time. And imagine on this day you decide to do everything just the way you want to. You wake up and get out of bed only when you want to. You eat and drink all and only what you desire. You perform the activities that make you happy. Perhaps you go and see your favorite sports team play, or you spend the day lying out in the sun reading a book, or you ride your bike out to a favorite spot. In a sense, what you do on this day is maximally free. It would seem there are no constraints on you forcing you to do things you don't want to do. Instead, you do exactly what you want. What Kane means by 'surface freedom' is exactly this ability to do what one wants, to satisfy one's desires. And, one might wonder, what could be missing? What more could there be to freedom of the will?

To show that there is a deeper kind of freedom, we must ask what it is in virtue of which you have these desires in the first place. Do these desires have their ultimate source in you, or do they arise from something else outside of your control? To see how that might be, consider a science fiction version of the previous scenario, where on that day indeed you do everything you want to do and these activities make you happy, but these desires and wants have all been programmed into you by some malicious neuroscientist. So the activities of the day really do give you what you want and really do make you happy, but they would not be what you want and these activities would not make you happy if you hadn't been programmed in just this way.

This suggests that surface freedom is not sufficient for genuine freedom of the will. Still, one might say, who cares, this is a science fiction case. In reality, I am not controlled by a malicious neuroscientist and my desires are my own.

But we may consider less science fiction versions of the scenario as well. Many philosophers have worried, and perhaps this is something you have worried about as well, that even those of us who are happy, those of

us who seem to think they have what they want out of life, lack freedom in an important sense that affects the character and value of their lives. For although there may be no evil neuroscientist controlling what I care about and what makes me happy, my desires and dispositions were shaped by factors beyond my control. The circumstances in which I was raised as a child, the norms of the society in which I live, the education I have had up until now, the books and websites I have read and movies I have watched, all of these things have constrained the type of person I am now, influenced what I desire and what it is that makes me happy. If I weren't subject to these constraints, would I still desire what I do, would these things still make me happy? It is hard to know, but it is also hard to avoid thinking there may be a deeper freedom of the will to be had. This deeper sense of freedom is what we mean when we talk about our being "the ultimate source" of our will and actions. We are now going to consider a reason to think it is lacking, even if somehow we could know we were not under the control of evil scientists or not shaped in some problematic way by our environments.

## 12.2 The Problem of Free Will and Determinism

When metaphysicians today worry about there being a threat to our free will, the main threat comes from the laws of the universe constraining our choices and actions. This comes from the **epistemic possibility** (possibility given what we know) that the basic laws of the universe are deterministic. This means that the laws are such that, given any complete state of the universe at a time and facts about what the laws are, it is possible in principle to deduce what the complete state of the universe will be at any later time. Given any past or present state of the universe, the laws determine what the future holds.

**Epistemic possibility:**  
something that is  
compatible with  
everything that one  
knows

How would determinism in the laws threaten free will? The rough idea is very simple. If determinism is true, then this means that the choices and decisions you make now, indeed everything you have ever done, was determined by states of the universe that occurred long before you were born and the laws. In fact, if determinism is true and there was, as cosmologists suggest, a Big Bang 13.7 billion years ago, it was already determined then what kind of person you would become and what you will be doing right now. None of these facts then about who you are and how you act, it would seem, are up to you. Given the laws and the facts about this early state of the universe, you couldn't have been any other way than how you actually are; you couldn't have done anything else. Or so it seems.

This reasoning was articulated very clearly in a famous argument by Peter van Inwagen. This is the Consequence Argument.<sup>2</sup> Its aim is to demonstrate the incompatibility of free will and determinism. We will follow

van Inwagen's way of presenting this argument. He starts by introducing a new sentential operator 'N.' 'Np' should be read as: p and no one has or ever had any choice about whether p. So, for example, if 'L' is a statement of the laws of nature of our universe, then 'NL' should be read as the statement that these laws obtain and no one has or ever had any choice about whether they obtain.

There are two inference rules that van Inwagen introduces and uses to run his argument. Both make use of this new sentential operator. Note that the first rule also uses the ' $\Box$ ' introduced in Chapter 10. Recall ' $\Box p$ ' should be read as 'Necessarily, p.' Here are the two inference rules:

- ( $\alpha$ ) From  $\Box p$ , one may infer Np.
- ( $\beta$ ) From Np and  $N(p \supset q)$ , one may infer Nq.

These are both very intuitive. If p is necessary, then p is true and no one has or ever has had a choice about whether p. And, if p and  $p \supset q$  are both true, and one never had any choice about whether they were true, then since by modus ponens, we may infer q, also q is true and one never had any choice about that either.<sup>3</sup>

Finally, let us again use 'L' to stand for the fundamental laws of nature. Take 'A' to be a statement about some action you actually perform. Any action will do. And finally take 'P' to be an exhaustive description of the state of the world at some time in the remote past long before you were born. We can now state the argument.

#### *The Consequence Argument*

- 1.  $\Box ((P \wedge L) \supset A)$  follows from the definition of determinism

Therefore,

- 2.  $\Box ((P \supset (L \supset A))$  follows from (1) and classical logic

Therefore,

- 3.  $N((P \supset (L \supset A))$  follows from (2) and ( $\alpha$ )
- 4. NP independent premise

Therefore,

- 5.  $N(L \supset A)$  follows from (3), (4), and ( $\beta$ )
- 6. NL independent premise

Therefore,

- 7. NA follows from (5), (6), and ( $\beta$ )

Stated in ordinary English, the conclusion of van Inwagen's argument says: For any action one has ever done or will ever do, although that action takes place, no one has or ever had any choice about whether it took place.

Assuming standard logic and the two inference rules ( $\alpha$ ) and ( $\beta$ ), van Inwagen's argument is valid. Moreover, it rests only on a few independent premises, all of which seem hard to deny. How could one deny (4): that no one has a choice about the state of the universe before they were born? Or (6): that they have no choice about what the laws of nature are? The only substantive and relatively controversial assumption is what one relies upon to get to (1): determinism. And yet the conclusion is very strong: it is that no one has or ever has had any choice about any action A they have ever performed.

What the Consequence Argument aims to show is that if the laws are deterministic, then given past states of the universe and these laws, no one has a choice about any action they have ever performed or will ever perform – where this means that no one is ever able to do anything other than what they actually do. If you ate cereal for breakfast this morning, you had to have cereal for breakfast this morning. You may have thought it was in your power to go out instead for a smoothie or Egg McMuffin, or to eat nothing at all, but this is wrong. Assuming the laws are deterministic, you had no choice. And thus, it would seem to follow, this, like all of the rest of your actions, was not really free; it was not really up to you.

### EXERCISE 12.1

#### The Consequence Argument

Consider each premise of this argument and each step leading from these premises or a minor or major conclusion. Identify the two most vulnerable parts of this argument and explain why they are so vulnerable.

Philosophers have not all been persuaded by this argument. But before we consider the main types of responses one encounters, we should try to see what one is committed to if one accepts the premises of this argument and follows van Inwagen where he thinks they lead.

There are two broad positions philosophers take today with respect to free will: those who think (as van Inwagen does) that free will is incompatible with determinism in the laws, and those who think free will is compatible with determinism. Those adopting the former position are called **incompatibilists**; those adopting the latter, **compatibilists**.

If one thinks that free will is incompatible with determinism, then one has another question to answer that is relevant to the question of whether we have free will: Is determinism true? And this marks another big division

**Incompatibilism:** the view that free will is incompatible with determinism

**Compatibilism:** the view that free will is compatible with determinism



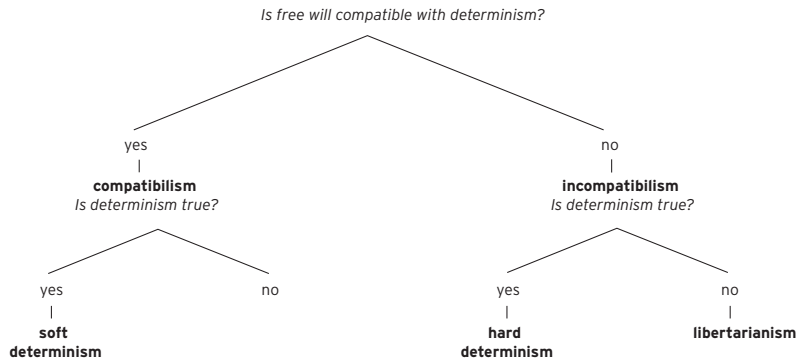


Figure 12.1 Main Views in the Free Will Debate

**Hard determinism:**

the view that free will is incompatible with determinism, and so, since determinism is true, we lack free will

**Soft determinism:** the view that determinism is true and it is compatible with the existence of free will

**Libertarianism:** the view that free will is incompatible with determinism and so, since we have free will, determinism is false

in the philosophy of free will. There are, on the one hand, those who think we have good reason to accept determinism. These philosophers end up being skeptics about free will. They are often called **hard determinists**. This name signifies that they accept determinism and think it has serious consequences for our sense that we are free. It contrasts with what is called **soft determinism**. This is the position of those philosophers who accept determinism but are compatibilists; that is, who don't think determinism is incompatible with our having free will. Finally, if you are an incompatibilist, there is another option: **libertarianism**.<sup>4</sup> Libertarians believe that free will is incompatible with determinism, but instead of rejecting the existence of free will, they reject determinism. Libertarians think, at least some of the time, that their actions are up to them, and so, as a consequence, they reject the claim that the laws of the universe are deterministic. We may represent these various positions as shown in Figure 12.1.

Since which position you adopt about free will depends so much on whether you think that the laws are deterministic, let's explore for a bit what reason there is to think determinism is true. What if anything suggests this and are we justified in believing it?

## 12.3 Determinism

Determinism is a view about the character of the laws of nature. It is a position that became motivated as scientific progress, especially after the development of Newton's mechanical framework, gave us reason to believe there was a set of fundamental, physical laws that govern the entire universe. The thought is that if we could just know where all of the fundamental particles in the universe are located and what their basic features are (their masses, velocities, and so on), we could use these laws to predict everything that will happen for all times into the future. In the early nineteenth century, Pierre-Simon Laplace (1749-1827) was so inspired by the ability of Newton's physics to accurately predict the behavior of physical

systems that he imagined the possibility of a supremely powerful intellect, one capable of using this system to predict all future behavior:

We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes. (Laplace 1951, p. 4)

This hypothetical intellect is now known as Laplace's Demon. If Laplace was this impressed with Newtonian mechanics, it seems he would be all the more impressed with the fundamental physical theories that have succeeded it. The special and general theories of relativity and quantum mechanics both have had empirical successes unrivaled by any previous physical theories. We know that neither is a final theory, as relativity and quantum theory will eventually need to be unified into something that encompasses both, but we can hope to eventually discover the fundamental laws. Perhaps once we know these and have the ability to describe an initial state of the universe, one could, given enough computing power, predict everything that will happen with perfect precision, as Laplace imagined.

Now there are some reasons, independent of the issue of free will, for thinking that perhaps determinism isn't correct. For, speaking of quantum mechanics, doesn't that theory tell us that the laws of nature are fundamentally indeterministic? And so even if we may be getting closer and closer to discovering the basic laws of the universe, this doesn't mean we are getting closer and closer to having the ability to predict what happens in the future with certainty.

There is a lot to say about the relationship between quantum mechanics and indeterminism. I will limit the discussion to two points, but first let's get just a bit clearer about what is meant by 'indeterminism.'

Like determinism, indeterminism is a position about the character of the laws. It is defined simply as the denial of determinism; so, it is the view that the laws are such that, given any complete state of the universe at a time and the facts about what the laws are, this *fails to* logically entail what the complete state of the universe will be at any later time. There are a number of reasons why indeterminism might be the case. Here are three possibilities. First, it could be because there aren't any laws of nature in the first place. It could be that the universe is such a disordered, jumbled place of random facts that there are no such things as laws in

any plausible sense of the word.<sup>5</sup> Second, it could be that there are laws of nature, but they do not cover all of the types of situations that may arise in our universe (perhaps they say how an object will behave when there are no forces acting on it, but do not describe what happens when there are forces). Or, perhaps there are laws of nature, but these are local laws that describe what happens in this or that part of the universe (here on Earth, or here in the part of the universe that we may observe), and there are other parts of the universe for which there are not even local laws that apply. Finally, and this is what many physicists and philosophers think follows from quantum theory, it could be that there are universal laws that govern what happens everywhere and in all types of situations, but these laws do not describe what will happen as a matter of necessity. Instead, they only describe what will happen as a matter of chance. Let's see whether it is true that quantum mechanics implies the laws are indeterministic in this sense.

The truth is that the interpretation of quantum mechanics is still widely debated even though the theory was developed a century ago, in the 1920s. The central dynamical laws of quantum mechanics, be they Schrödinger's equation or the relativistic variants thereof, are all purely deterministic laws; they do not make any reference to chances. But, some think that Schrödinger's equation cannot be all there is to quantum mechanics and that there must be some additional laws describing the chances that there will be a so-called collapse of the wave function.<sup>6</sup> If this is correct and there must be a probabilistic law governing wave function collapse, then indeed the laws of the universe will be indeterministic. However, this is by no means a settled opinion. One popular interpretation of quantum mechanics that many physicists and philosophers take seriously, the many worlds interpretation, denies that quantum mechanics needs supplementation with any indeterministic law.

Despite the fact that it is contentious whether the fundamental physical laws are deterministic or indeterministic, some have argued that even if the laws are indeterministic, this still doesn't help one avoid the kind of determinism that threatens our having free will. The reason is that the probabilities one finds in the laws of quantum physics have important consequences for the behavior of subatomic matter and radiation, but they have negligible consequences when it concerns the behavior of objects at the scale of the macroscopic.<sup>7</sup> When we consider systems of particles as large as human beings, indeed even as large as the cells making up human beings, the probabilities in the quantum laws "wash out," and it becomes possible to tell for most systems how they will behave at the next time. Strictly speaking, this isn't quite right. If the quantum laws are truly indeterministic, there is always a tiny chance even for a large system the size of (say) a human being that we will observe some unexpected quantum effect. However, the chances of these effects are so tiny, they

are so unlikely, that such effects would not suffice to give us the freedom and control we want over the choices we make every day. Or at least this is a common view. In the later section on libertarianism, we will see how some philosophers have tried to exploit quantum indeterminism, however rarely it may have a noticeable significance in the lives of human beings, to save free will.

## 12.4 Compatibilism

So determinism, even if it is not a settled feature of physics, is a live option and seems to at least obtain at the macroscopic level of human beings and the medium- and large-sized objects with which they interact. If this is right, and we follow van Inwagen's argument where it leads, then it looks like we as people never have any choice over our decisions and actions. We are never free, since everything we think and do is determined by events that took place long before we were born and the laws of nature. But should we follow van Inwagen's argument where it leads? A large number of metaphysicians think we should not – these are the compatibilists about free will – and they have a variety of ways of responding to the Consequence Argument. We will just discuss two such responses here and the impact these responses have for how we should understand free will.

One response some have tried is to press on premise (6) and whether it is indeed the case that no one has or ever had any choice about the laws of nature. The premise certainly seems intuitive, but David Lewis, in his paper “Are We Free to Break the Laws?” argues that it is ambiguous in a way that calls into question the soundness of van Inwagen's argument. To see this, let's hold determinism fixed in the background and just ask what it takes for it to be true that you could change a fact about whether one of your present actions occurs. Earlier, we considered the case of your having breakfast this morning. Suppose again that you had cereal. Now ask: at the time when you were about to have the cereal (let's call that time 't'), could you have made it so that you did not have the cereal? This is Lewis's question: what would make it the case that you did not have that cereal this morning? And is it possible that you didn't have the cereal this morning?

Now in any usual sense of the word 'possible,' of course this is possible.<sup>8</sup> It is of course a *possibility* that you did not have cereal this morning, but had (say) a smoothie instead. There is no logical contradiction in the idea of your having a smoothie. But you may ask, is it a possibility that you did not have cereal this morning and everything about the world was the same up until the time when you made the decision? Again, Lewis will say “yes”: there is certainly no contradiction either in your past history being just the same up until time t and then instead of choosing cereal, you choosing a smoothie.

Now what is true is that in a situation like that, something would have to give, if determinism were true (which, remember, is the background assumption we are making), then things couldn't have been just the same up until *t* *and* the laws were the same as those at our world *and* yet you chose a smoothie. But this is not the question. The question is whether there is a possible world in which things are just the same up until *t* and yet you have a smoothie. And this is possible. What we will have to say is that at that world, there are different laws of nature. But this doesn't mean that anything outrageous happens, that you somehow have special powers to break the laws of nature at this world. So, Lewis notes, the ambiguity in premise (6) is between the following weak and strong claims:

*Weak claim:* I have the ability to do something such that, if I did it, a law would be different.

*Strong claim:* I have the ability to break a law.

Lewis's point is that since you do have the ability at time *t* to have a smoothie, the weak claim is true. The world in which you choose the smoothie is just a world at which the laws are different. But this doesn't mean you have supernatural abilities in any sense.

At this point, there is a disagreement that was never resolved between the incompatibilist van Inwagen and the compatibilist Lewis. The incompatibilist in general will think that Lewis is misunderstanding what the issue is in the argument. For him, the issue is: given determinism, and how things are up until *t*, and the *actual* laws, we fail to have any choice about whether we have cereal. For Lewis, to understand our freedom, we have to understand that part of what is involved in free action is that in a sense we have the ability to do otherwise. This doesn't mean we can create exceptions in the actual laws of nature, but rather in the sense of acting such that the laws of the world are different than what they otherwise would have been. Probably a large part of this disagreement stems from a disagreement about the metaphysics of laws. Lewis thinks that laws of nature are just more or less descriptions of what actually happens at a universe (see Text Box 12.1). And so if people act differently at a world, this will entail that the laws at that world are different. Van Inwagen seems to be assuming in the background a different theory of laws, where they could not be ultimately dependent on what people choose to do.<sup>9</sup>

#### TEXT BOX 12.1

##### **Humeanism about the Laws**

We have now seen the laws of nature play a role in several central debates in metaphysics. So it is worth briefly outlining a couple of theories about the nature of these laws. One

common view about the laws of nature is what is now referred to as **Humeanism about laws**. Humeans believe that the facts about what the laws are are ultimately explainable in terms of or grounded in facts about what happens at a world, that is, facts about what kinds of objects and events there are and how they are distributed over spacetime. Anti-Humeans think that the facts about what the laws are are not explainable in terms of or grounded in facts about what happens. Rather, the facts about what the laws are are additional facts over and above what happens at a world. The facts about the laws instead *explain* what happens.

Humeanism is named after David Hume because it was he who held there were no necessary connections between distinct existents. Since if laws were fundamental, this would mean there are fundamental, necessary connections between the events that take place; the Humean wants to explain what appear to be necessary connections in terms of more basic facts about what happens just as a matter of fact (not as a matter of necessity).

Lewis sketches his particular Humean theory of laws in his paper "New Work for a Theory of Universals." Some influential critiques of Humeanism in its various versions, as well as defenses of rival anti-Humean positions, may be found in David Armstrong's *What is a Law of Nature?* and Tim Maudlin's *The Metaphysics Within Physics*.

So this is one issue compatibilists have raised for the Consequence Argument. Perhaps a more frequent criticism that really gets to the heart of the issue of what is involved in freedom of the will is directed at the inference from premises (5) and (6) to the conclusion (7):

5.  $N(L \supset A)$

6.  $NL$

Therefore,

7.  $NA$

Recall that this inference relied on van Inwagen's rule ( $\beta$ ):

( $\beta$ ) From  $Np$  and  $N(p \supset q)$ , one may infer  $Nq$ .

Compatibilists might argue that even though it is true that one has no choice about what happened before one was born or what the laws are, and even though past events and the laws may determine one's actions, it doesn't thereby follow that one doesn't have the choice about the actions one undertakes. Whether one has a choice about the actions one undertakes depends on the causal history of these actions, what sorts of earlier states they were caused by. So, rule ( $\beta$ ) is not universally valid.<sup>10</sup>

This is a core insight of compatibilists: even if our actions are in a genuine sense determined by prior states of the universe and the laws of nature, they can still be up to us, if they are caused in the right way by our prior mental states. For example, as the philosopher A.J. Ayer (1969) noted, there is an important distinction one may draw between those

actions we undertake under constraint and those for which we are not constrained. It is plausibly those actions that are not constrained that are the ones that are up to us in the sense that matters for free will.

What does constraint amount to? Ayer gives several examples to show what he has in mind. Someone may constrain your action by putting a gun to your head, or by gaining a "habitual ascendancy" over you so that, without force, you do what they want. One may also be constrained by one's own uncontrollable urges. The kleptomaniac is impelled to steal and feels satisfied when she does, but these are not free actions. They are not ones that are under her control. What seems common to these cases is that although the agent is doing something, she is not doing what she ultimately wants. For Ayer, for an action to be free, it must be unconstrained.

The compatibilist thinks we can have free will even if our actions are the results of earlier states of the universe and the laws. For the compatibilist, determinism doesn't stand in the way of freedom, because what is really important to our concept of freedom is that our actions come from us in the right way. In doing so, many compatibilists adopt what is called a **sourcehood conception of freedom** rather than a **leeway or alternative possibilities conception**. Freedom doesn't require there being a leeway or opening given past states of the universe and the laws for me to act one way as opposed to another. It just requires me to be the proper source of my actions.<sup>11</sup> Probably the simplest compatibilist view along these lines was proposed by Thomas Hobbes (1588-1679) in his *Leviathan*. For Hobbes, an action is free if it is in line with what a person has the desire or will to do (Hobbes 1651, Part II, Chapter 21).

In fact, many compatibilists, Ayer included, argue not only that there isn't tension with determinism in the laws, but that our sense of freedom importantly relies upon the laws being deterministic. For if the laws were *not* deterministic, then there would be no guarantee, given what one wants and what one chooses, that these choices would ultimately allow one to have any control over what happens. We rely on determinism in the laws in order to be able to predict which sorts of behavior will lead to which sorts of consequences.

I call Hobbes's a "simple" form of compatibilism and it is. Recalling our earlier discussion of the distinction between surface freedom and deeper freedom of the will, one might worry that this view doesn't build in enough to capture the sense of freedom we most care about. To say that a given action is unconstrained, where this means it is an action that is in line with what we desire, is compatible with its not having its ultimate source in one's true character. One might propose to the compatibilist that she should adjust her view to call free only those actions in line with what a more self-reflective, rational person would want. For example, say I am the

**Sourcehood conception of freedom:** a conception of freedom according to which it requires an agent to be the proper causal source of their actions

**Leeway or alternative possibilities conception of freedom:** a conception of freedom according to which it requires the ability to do otherwise

kind of person who desires all of the time to sleep and watch television. These are, as a matter of fact, the actual desires I have, but I also recognize that if I were the person I really wanted to be, there would be other things I would want to do and so these would not be my desires anymore. One might strive for an account of freedom that doesn't count as free those actions that are simply consistent with one's desires, or that stem from one's desires, but those that are consistent with some deeper fact about one's psychology or moral standards.

There are many ways of modifying Hobbes's simple compatibilist account of what freedom is in order to accommodate a deeper sense of freedom. One suggestion made by Harry Frankfurt (1971) is to shift from talk of ordinary, first-order desires (desires to do some action A), to talk of second-order desires (desires to desire some action A). I may very well desire to sleep and watch television all day. But this may not be what I desire to desire, and if so, one might think, I am only really free if my actions line up with my higher-order desires. Other compatibilists have argued that instead of using desire in an account of free action, one should appeal to the use of reason on the part of the agent in choosing an action. One might worry that in order to really rule out the kleptomaniac's actions as free, one cannot allow that free actions are those that are merely in line with one's desires or what one wants. The kleptomaniac may really want to steal. Instead, what is distinctive about actions that are genuinely free is that they are the types of actions that stem from an agent's rational assessment of an action and choosing what to do based on reasons in favor of that action. This type of reasons-responsive compatibilism has been developed by the philosopher John Martin Fischer (1994). Finally, some compatibilists have argued that free actions need to be tied to the agent's moral standards. Susan Wolf (1990) sees free will as a matter of a person's attending correctly to what is right and good.

## EXERCISE 12.2

### Versions of Compatibilism

Consider one action of significance to you in your life, an action that you take pride in and would consider to be a free action. What is the most promising compatibilist account of why this action was free? Which, of the compatibilist accounts mentioned, seems best to capture the sense in which this action was "up to you" and not constrained? Despite this being the best of the compatibilist accounts, does this, in your view, still leave something out?



However the compatibilist develops their position, the key insight is that there is a distinction among all of the actions there are between those that are caused in the right way by earlier mental states of the agent (by being the results of the agent's desires, higher-order desires, reasons, or moral standards) and those that are not. And this, for the compatibilist, tracks the distinction between those actions that are free and those that are not, even if all events whatsoever are determined by earlier states of the universe and the laws.

## 12.5 Frankfurt Cases

In the previous section, we discussed two ways in which compatibilists have responded to the Consequence Argument. However, an incompatibilist might argue that even if it is the case that (a) there are logically possible worlds in which an agent acts differently and there are different laws of nature and (b) we may draw distinctions between actions with different kinds of causes, still it does not change the fact that if determinism is true, all of my actions are necessitated by earlier states and the laws. So given earlier states and the laws, no one could ever have acted in any other way than they actually did. And so no one is ever free in the leeway or alternative possibilities sense of the term. The incompatibilist can then argue that free will requires this power to do otherwise, given earlier states and the laws.

In response to this claim, compatibilists have conjured interesting thought experiments showing us that we have no trouble at all imagining cases where someone makes a decision or acts in a way that is perfectly free and yet it turns out that they were incapable of deciding or acting otherwise. One old example comes from the philosopher John Locke (1632-1704) who imagined a situation in which a man wakes up in a room that is, unbeknownst to him, locked from the outside. The man decides on his own to stay in the room. The decision and action are both free, and yet the man does not have the ability to do otherwise.

The philosopher Harry Frankfurt, in a paper in 1969, constructed an alternative case that has seemed to many philosophers to be a more compelling illustration of the fact that agents can act freely even if they do not have the ability to do otherwise.<sup>12</sup> This sort of case is now widely referred to as a **Frankfurt case**. Frankfurt cases involve the following features:

- There is an agent who makes a decision at some time *t* to engage in some action *A* and then acts on this decision without interference from the outside.
- Unbeknownst to this agent, there is an outside party with an interest in the agent's doing *A*. The outside party has planted a device in the agent's head ensuring that if the agent does not choose on their own

**Frankfurt case:** a case in which intuitively one acts freely and so is morally responsible for an action, and yet one did not have the ability to do otherwise

to do A by  $t$ , the device will forcibly set the agent's body in motion to bring about A anyway. If the agent does decide to do A by  $t$ , then the device does nothing.

For example, suppose there is an assassin Alice who is hired by the mob to murder an important political figure Beatrice. The mob boss who hired Alice, Carlo, finds it very important that the assassination is carried out and does not entirely trust Alice, so Carlo has a device planted in Alice's head while she is unconscious and without Alice's knowledge. The purpose of the device is to ensure that if she should back out of the assassination, the device will make it so that Alice undertakes the assassination anyway. As it happens, Alice carries out the assassination as planned. Of her own accord, she makes the decision to murder Beatrice, the murder takes place, and the secret, implanted device never needs to interfere with Alice's body. Many find this a good example of a free action, although because of the presence of the device, it appears to be a case in which the agent is not able (does not have the "leeway") to act otherwise.

What is important and clever about this case, like Locke's, is the fact that even if there are constraints that may have prevented the agent from doing something other than what she chose, those constraints turn out not to be operative. In the actual circumstances they do nothing to influence the agent's action. We have two clear cases then in which an agent acts directly on the basis of their own decisions. It is thus puzzling why the mere possibility of the agent's action being constrained should affect whether in the situation under consideration, the act (and the decision too) was free. The Frankfurt case thus seems to speak in favor of a sourcehood, rather than a leeway or alternative possibilities, conception of freedom.

### EXERCISE 12.3

#### Frankfurt Cases

How might one respond to the Frankfurt case just described in order to defend the principle that free will requires the ability to do otherwise?

## 12.6 Manipulation Arguments

More recently, incompatibilists have developed another kind of argument to defend their position, one that relies on different intuitions than the Consequence Argument, and thus requires a different kind of response. This kind of argument is called the Manipulation Argument

because it relies on the consideration of scenarios in which an agent is manipulated to act in a certain way, and so intuitively, is no longer the ultimate source of their actions. The incompatibilist uses such scenarios to argue that compatibilist accounts of freedom fail. They fail because agents can satisfy the requirements of compatibilists accounts and yet at the same time be manipulated to act, so that, intuitively, their actions are not free.

Most famously, Derk Pereboom (2001, 2014) introduced a series of four cases intended to undermine the various compatibilist accounts of free will discussed at the end of Section 12.4. The cases share the following form. We imagine an agent (Professor Plum) murders a Mrs. White. Pereboom sets the case up so that Plum's action satisfies whatever conditions a compatibilist would claim is required for free will. For example, Plum's action satisfies the requirements of Hobbes's simple compatibilist view, because Plum acts in line with his desires. It satisfies the conditions of Frankfurt's higher-order view, because Plum's action is in line with both his first- and second-order desires. His action also satisfies the requirements of Wolf's view, because his action follows from his determination of what is right and good. And so on, for any compatibilist account of what it takes for an action to be free.

In all four cases, Plum murders White and he satisfies all conditions of the compatibilist accounts. However, the cases vary in how and whether Plum is manipulated by external agents. Here are Pereboom's cases:

*Case 1:* A team of neuroscientists have the ability to manipulate Plum's brain states by radio technology. In this case, they intervene on Plum just before he begins to decide to kill White. They produce neural states in him that realize the desire to kill White, the higher order desire to want to desire to kill White, the belief that killing White would be right, etc. Plum is then causally determined to kill White as a result of this reasoning.

*Case 2:* A team of neuroscientists programmed Plum at the beginning of his life so that later, at this exact time, he has neural states that realize the desire to kill White, the higher order desire to want to desire to kill White, the belief that killing White would be right, etc. Besides this programming at birth, Plum is otherwise a normal person. Plum is causally determined to kill White as a result of this reasoning.

*Case 3:* Plum grows up and lives in a community that engages in training practices that will result in him having neural states that realize the desire to kill White, the higher order desire to want to desire to kill White, the belief that killing White would be right, etc. Besides these community training practices, Plum is otherwise a normal person. Plum is causally determined to kill White as a result of this reasoning.

Case 4: Plum is a normal human being, living in normal circumstances. The past states of the universe and laws of nature causally determine that he has neural states that realize the desire to kill White, the higher order desire to want to desire to kill White, the belief that killing White would be right, etc. Plum is causally determined to kill White as a result of this reasoning.

The fourth case is supposed to be an ordinary case of how agents act in a deterministic world. Agents form mental states. They deliberate over which actions would be best, and act accordingly. At the same time, all of their actions are determined by earlier states of the universe and the laws of nature.

Pereboom expects everyone to share the intuition that Plum is not free in the first case. His murder of White was not free because it was not up to him. It was an action chosen by the neuroscientists that have manipulated him. However, Pereboom notes that there is no relevant difference between this case and the fourth, ordinary case that could bear on whether the action is free, according to compatibilist accounts of free will. In both cases, Plum has deliberated and chosen his action in line with his desires, reasons, and moral evaluations. So, if he isn't free in the first case, then he can't be free in the fourth case.

The Manipulation Argument may thus be summarized in the following way (see Pereboom 2022, pp. 33-34):

*The Manipulation Argument*

1. If an agent is causally determined to act by neuroscientists who manipulate his brain, then that agent's action is not free, even if it satisfies the conditions of a compatibilist account.
2. There is no difference that is relevant to freedom between the actions of such a manipulated agent and the corresponding actions of an ordinary, causally determined agent.

Therefore,

3. The actions of ordinary, causally determined agents are not free.

There are two premises to this argument. The first is a statement of the intuition that Professor Plum in the first case is not free. The second expresses the fact crucial to the setup of the cases that, by the compatibilist's own lights, there is no difference between Professor Plum in the first and last cases that could be relevant to whether his action is free.

Compatibilists have challenged both premises of the Manipulation Argument. Some, like Michael McKenna (2008) argue that they don't clearly have the intuition that Plum is not free in Case 1. McKenna asks, if it is truly the case that the Plum with the team of neuroscientists ("Team Plum") really satisfies all of the compatibilist conditions for freedom, so

that he is acting rationally in line with his desires and moral standards, then why can't we consider him to be free? More generally, McKenna questions whether we should take manipulation to be a barrier to personal freedom, pointing out that many of our actions, including the particularly significant ones that make us who we are, are the result of external manipulation that wasn't under our control. An unplanned pregnancy or unexpected religious experience can make us change the kind of actions we undertake going forward and the people we choose to be. McKenna argues "here again, in these sorts of cases, cases that we do encounter in ordinary life, we have what look like more mundane cases of manipulation" (2008, p. 156). And, the manipulation does not bar the freedom of the resulting actions.

Other compatibilists (e.g., Fischer and Ravizza 1998) have responded to the Manipulation Argument by challenging its second premise. They argue that there will inevitably be a difference between the first and last cases that is relevant to whether Plum's action is free. In Case 1, the neuroscientists are intervening in Plum's thought processes. Thus, "Plum's decision," is not *his* decision at all. The mental states that result from the neuroscientists' manipulation are not the result of Plum's earlier personal history. And this is why Plum is free in the last case, but not in the first.

Pereboom's four case version of the Manipulation Argument is only one version of the argument. Another influential version, offered by Alfred Mele (2006) is called the Zygote Argument. Mele asks us to imagine an agent Ernie who is manipulated as a zygote by a goddess (Diana). Diana manipulates the zygote so that when Ernie is thirty years old, he will deliberate and choose to carry out a particular action A. Here again, we are supposed to have the intuition that, regardless of whether the agent satisfies the compatibilist conditions on freedom, they cannot be a free agent. After all, their actions up to this age were all manipulated so as to lead up to the action A thirty years later, so how could A be free? Mele himself does not take this argument to show that compatibilist accounts of freedom fail, because he is not certain whether Ernie is free or he isn't. However, he produces an interesting variation on the Manipulation Argument, by considering a scenario in which an agent has a personal history that is uninterrupted from a time before his birth.

Metaphysicians continue to debate the soundness of manipulation arguments like those of Pereboom and Mele. As in the case of the Consequence Argument, incompatibilists will tend to defend the arguments as sound; compatibilists will work to raise challenges to the arguments' key premises or inferences. Many think that manipulation arguments provide a better challenge to the compatibilist by not relying on a leeway or alternative possibilities conception of freedom, and instead taking on board the assumption that freedom requires being the source of one's actions.

**EXERCISE 12.4****The Zygote Argument**

State the zygote argument in numbered premise form. How could a compatibilist most plausibly argue this argument is unsound?

**12.7 Libertarianism**

There are a good many compatibilists today, those who think that we have free will even if the laws are deterministic. But there aren't enough compatibilists to say by any means that it is the orthodox view in philosophy on the topic of free will.<sup>13</sup> Just as there are a good many compatibilists, there are also very many philosophers who endorse libertarianism. These libertarians do not disagree with the compatibilist that we possess free will, but they do disagree about what is required for us to have it. The libertarian thinks it is not merely enough that our actions stem from our desires, or our higher order desires, or our goals or reasons for those actions to be free. For our actions to be free, our actions must not be determined by the laws. And so, the libertarian holds, determinism must be false.

There is an immediate problem that arises if one rejects determinism. If determinism is false, then this means that given that the world is in a given state at a one time, this does not guarantee that the world will be another determinate way at a later time. And so, one might form a specific intention, make a decision to act in a certain way, but if the laws are indeterministic, then one can't thereby guarantee that one's actions will follow in the way one expects them to from these intentions and decisions. The connection between these psychological states and your later behavior and the eventual state of the world that results would seem to be a matter outside of your control. If determinism is false, your actions will not be determined by the laws, yes. But they will not be determined by you either.

This is the main challenge for the libertarian: to explain how your actions may be up to you in the sense that matters for free will if the laws are indeterministic - why your actions are not then just arbitrary or a matter of chance or random or uncontrolled. We will discuss two ways that libertarians have tried to meet this challenge.

This first is a form of libertarianism that received an influential defense by the metaphysician Roderick Chisholm (1916-1999). This view is called **agent causal libertarianism** or the **agent causation** view. The main idea behind the view is that your actions may be free and not the result of chance or random when they are directly caused by you. Chisholm thought

**Agent causal libertarianism:** the view that human agents are sometimes causes and that is how their actions can be free

that if your actions were the results of any earlier event or states, even your own psychological states, these actions couldn't truly be yours since then in a sense you wouldn't have ultimate control over them. Once the desire or intention was formed, the action would follow automatically. This is what makes Chisholm an incompatibilist. But he also recognized that it would not do to say that free actions are caused by nothing; that they just arise out of nowhere. In that case it would seem even more obvious that our actions aren't free. The alternative is to say that there is a cause of any free action, but this cause isn't any earlier event or state, even a psychological state. The cause for a free action must lie outside of the causal sequence of events that is governed by the laws of nature. Free actions involve agents acting in a way that transcends the causal chain of events to directly bring about changes in the world.

To some philosophers, Chisholm's reasoning and the agent causal view have been very attractive, even obvious. However, to many it is mysterious. One may ask: What are these agents that somehow lie outside of the order of events in the causal sequence that is the result of the laws? How do they do what they do without exploiting and acting in terms of the laws that govern our universe? It is probably not surprising to hear that Chisholm's position is one that is especially popular among mind-body dualists, those like Descartes who think that there is a fundamental distinction between mental substances (minds, souls) and physical substances (bodies, matter) and that we, as persons, are to be identified with mental substances not physical substances. If one views the agent in this way, as a nonphysical substance, then it doesn't seem much to also take on board the idea that the agent acts in a way that may transcend the law-governed chain of events.<sup>14</sup>

This doesn't mean one has to reject physicalism in order to be a libertarian about free will. An alternative position, event causal libertarianism, tries to make sense of freedom in an indeterministic world while holding onto the view that the only kinds of entities that may be causes and effects are events or states of some kind. There is a fundamental metaphysical disagreement then between these two kinds of libertarian over what are the relata of causal relations (see Text Box 11.1). The event causal libertarian thinks causation always only links events or states of the world, whereas the agent causal libertarian thinks at least sometimes certain kinds of substances can be causes, in particular, agents in cases of free action.

There are many versions of event causal libertarianism available. To give a flavor of what some of these views look like, I will just describe one version that has been much discussed in the free will literature. This is Kane's (1996) account. Kane exploits the indeterminism that is thought by some to be a consequence of quantum mechanics to show how some of our actions may

be free. His account centers around the discussion of a class of important, life-directing actions, what he calls **self-forming actions**.

Kane doesn't think it is necessary or plausible for the libertarian to try to argue that the majority of the actions in our daily lives are free in the deep sense that matters to us for free will. Every day, we do an uncountable number of things, often without thinking very hard about what we are doing. We do things on whims, or out of habit, or out of a sense of duty or a desire to stick to plans we've made or settled on in the past. I wake up, brush my teeth, have breakfast. I practice my instrument, do the reading I've assigned for class, or go to the gym. Some of these actions are important, many are constitutive of the kind of person I am or who I want to be. But, Kane would argue, most of them are not free in any deep sense. And thus it is fine, Kane can concede, that these are actions that are caused by earlier states of the universe (perhaps including earlier psychological states) and determined by the laws.

It is, however, important for Kane, as it is generally for libertarians, that there are some actions that are free in a deep sense and so are not necessitated by earlier states of the universe and the laws. These are those actions that make one the kind of person that one is; these are the self-forming actions. To illustrate the kind of action Kane has in mind, we may consider a famous case described by Jean-Paul Sartre (1905-1980) in his discussion of the self and free action (1946/2007). Sartre was writing in German-occupied France during World War II.<sup>15</sup> He describes the case of a young student who was involved with the French Resistance movement and wanted very much to join the effort to free his country by fighting. This would require him to be far away from his family. What made this decision especially hard was that at the time, his mother was very sick and required him to take care of her. If the student left to go fight with the Resistance, his mother would certainly become worse and might die. What Sartre was most interested in demonstrating with this example is the fact that although there are most certainly facts about the kinds of people we are, facts, he says, about our *essences*, these are facts that are not determined by something external to us, something set in place before we are born or something transcendent like a God or innate nature. We are responsible for our essences, for the kinds of people we are. In fact, this is how Sartre defined the view for which he is most famous: **existentialism**. This, Sartre characterizes as the view that existence precedes essence rather than the other way around. It is not that there are facts about the kinds of people we are and then we act in the way we do because it was already determined we are that way. (This would be for our essence to precede our existence.) Rather, we act in a certain way and this makes us the kind of people we are. In the case of the student, he decides to fight for the Resistance and thus becomes this kind of person, the kind of person

**Self-forming actions:** important actions in the life of a person that decide the kind of person they will be

**Existentialism:** the view that it is the kind of things we do that determines our essences, the kind of people we are; we do not possess innate essences that determine who we are and what we will do



who would put his political causes above the strong pull of family. It is the student's action that makes him the kind of person he is, not the other way around.

These are the types of actions in which Kane is most interested. Perhaps you have never had to choose between taking care of your sick parent and going off to fight in a war, but Kane (and Sartre too) would argue that at certain points in your life, you have had to make important decisions, and you are torn about which is the correct path to take. There were two or more paths you could have taken and these would have led to your becoming a different sort of person. You choose one and this then forms you, forms the kind of person you are, and thus leads to many of the more mundane actions you perform later as a result. If you have never performed any of these self-forming actions, then you are not truly free in the sense that should matter to us as persons.

We may now better understand the sort of actions on which Kane thinks an account of free will should focus, but this doesn't yet say anything about how he proposes to solve the libertarian challenge of reconciling freedom with indeterminism. His account has two main parts. First, Kane must articulate in what respect these self-forming actions are not determined by earlier states of the universe and the laws. Second, Kane must explain how actions that are not so determined may nonetheless not be random or arbitrary, but free actions of a rational agent.

On the first issue, Kane points to features that are special about those situations in which agents are making self-forming actions. In these cases, Kane speculates that there may be two or more processes taking place in the brain. In the case of Sartre's student, there may be a well of neural activity that under normal circumstances would lead to the decision and action of going to war and another that under normal circumstances would lead to the student's staying at home and caring for his mother. Since these neural processes lead to incompatible behavior, they can't both ultimately lead to action. One must, as it were, win out. Here is where indeterminism enters the picture. Kane speculates that in cases like this, activity in the brain may be described in the technical sense of the scientist as chaotic. One thing that is characteristic of chaotic processes is their unpredictability. Often their unpredictability can be traced to their turning on a very small detail.<sup>16</sup> This may be a tiny, indeterministic quantum process. In the case of a self-forming action, things could have, given earlier states of the agent's brain, gone one way or the other. The fact that this is an indeterministic process though ensures that whichever action results was not determined in the sense that would be a threat to free will.

What remains then is only for Kane to explain how it could be that such an undetermined act could be free. If the action was not *caused*

by earlier mental states of the agent, but things could have gone either way, then in what sense could it truly be free? And here Kane points to the nature of these special cases of self-forming actions. It is not that any action that results from quantum indeterminism is free, but, in these cases, we can see that there are only a small number of actions that could have resulted. Either, if they had obtained, would have had their source in the agent, because both stemmed from these sorts of self-constituting activities.

There is a lot to say about both of these models of libertarian free action: Chisholm's agent causal libertarianism and Kane's event causal version. And, indeed, many objections have been raised to both accounts. But if one wants to understand how at least some of our actions can be free and one finds freedom incompatible with determinism, one has to develop some account. This is a subtle and difficult project. It has led some incompatibilists, van Inwagen (2000b) most notably, to claim that free will is and will always remain a mystery. It is extremely difficult to see how freedom could be compatible with determinism, but it may be equally difficult to see how it could be compatible with indeterminism as well.

## 12.8 Skepticism about Free Will

And this brings us finally to the last group of views one finds about free will. This is to deny the phenomenon altogether. We have already described a view called 'hard determinism.' Recall: this is a version of incompatibilism according to which determinism is true and so, free will does not exist. A stronger position, one that is compatible with hard determinism, is a position that Pereboom has labeled **hard incompatibilism**. This is the view that yes, free will is incompatible with determinism, but it is also incompatible with indeterminism, because in an indeterministic world, everything that happens is arbitrary, or a matter of chance, and so beyond anyone's control. The hard incompatibilist doesn't need to have a view about whether the laws of nature are deterministic or indeterministic. They will say that, either way, free will couldn't exist.

**Hard incompatibilism:**  
the view that free will is incompatible with both determinism and indeterminism and so we are not free

Some who have thought seriously about this issue are not very bothered by this. For example, the psychologist Daniel Wegner (2003) argues that we lack free will, but that this is something we can live with because all that matters to our lives is that we as humans maintain the illusion that we are free agents. The illusion is sufficient to keep us engaged and see our lives are meaningful. The neuroscientist Sam Harris (2012) also defends this position, although he questions the utility of the illusion, that realizing we are "biochemical puppets" can allow us to be more aware of the true causes of our behavior and perhaps improve our lives.

**TEXT BOX 12.2****The Libet Experiment**

In the 1980s, the neuroscientist Benjamin Libet and his collaborators conducted the first of a series of experiments that many argue should make us believe we lack free will. Libet had his experimental subjects wired with electrodes that measured their brain activity while they performed a simple action, for example, moving their finger or pushing a button. They were told they could perform the action at any time they wished, and should make a note of the time when they formed the conscious intention to do so. Libet showed that at a time about 550 milliseconds prior to the conscious intention, there was a swell of activity in the subjects' brains, what neuroscientists called the "readiness potential" (Libet et al. 1983). Skeptics about free will have used the results of Libet's experiments to argue that free will is an illusion. As Harris puts it:

One fact now seems indisputable: Some moments before you are aware of what you will do next – a time in which you subjectively appear to have complete freedom to behave however you please – your brain has already determined what you will do. You then become conscious of this "decision" and believe that you are in the process of making it. (2012, p. 9)

A premise in the skeptic's reasoning is that for your action to be free, it must be caused by your conscious intention to act. However, conscious intentions occur too late to be the causes of actions, since they occur only after the readiness potential. So, it seems to follow, human actions are never free.

Both compatibilists and libertarians have had much to say about neuroscientific experiments like that of Libet. One natural reply to the skeptic's argument is to question the assumption that the event preceding the conscious intention is not in an important sense the actor's decision. Some have questioned the accuracy of the subjects' assessment of when the conscious intention took place (e.g., Dennett 2003). Since the 1980s, many variations on the original experiment have been conducted and are compatible with the initial results. Philosophers and scientists continue to debate the lessons we should draw from these experiments about free will.

However, one might worry that a lot is lost if we are forced to give up seeing ourselves and others as free agents. The biggest issue turns on the fact that it is almost unanimously agreed by philosophers that there is a necessary connection between free will and moral responsibility. One can only view oneself or someone else as morally responsible for their actions if those actions are up to them. If someone's actions aren't their own, then it is not reasonable to blame that person (if one judges these actions to be morally wrong) or praise the person either (if one judges these actions to be good).

It is not an esoteric philosophical point that free will is tied up with moral responsibility. We are familiar from everyday interactions, for example,

with small children or the sick, that sometimes someone does something we don't like, but it isn't right to blame that person since the act wasn't really in an important sense up to them. They couldn't have helped what they did. Similarly, we all know that sometimes we do good things, but not because we wanted to or because these are actions were the result of our free choices, and that when this happens, these are actions we don't really deserve to be held morally responsible for. We don't deserve to be praised for them. If you make the effort to phone up your grandfather every week, but not because you really want to, but because your parents threaten to remove your monthly allowance if you don't, then you hardly deserve praise for being the faithful grandchild.

In his work, Pereboom (2001, 2014, 2021) acknowledges the many radical consequences the loss of moral responsibility would seem to have for the way we live our lives and interact with others. If it were true that none of our actions were up to us and so we uniformly lacked moral responsibility for them, this would mean that we couldn't reasonably praise or blame anyone for anything they ever did. You wouldn't have reason to take pride in the successes in your life. You couldn't reasonably blame the thief, murderer, or rapist. You couldn't see the important relationships in your life, your friendships and loves, as stemming from another person's free decision to spend time with you and make sacrifices on your behalf.

This might cause some to rethink the denial of free will in the first place. This, after all, is a lot to give up. But, Pereboom argues, as surprising and challenging as it is to see ourselves in this way, free will is something that must be given up since it is neither compatible with determinism nor indeterminism and these are the only two metaphysical options there are. (Indeterminism is after all just the denial of determinism!) The challenge then is to explain how we might continue to see our lives and relationships as valuable even without the existence of free will; and how we might justify locking people up in jail and imposing other penalties, even if people are not actually morally responsible for any of their actions. This is, ultimately, the main challenge for the metaphysician or scientist who holds a skeptical view about free will and moral responsibility.

### **EXERCISE 12.5**

#### **Compatibilism, Libertarianism, Hard Determinism, or Hard Incompatibilism?**

We have now seen four major views on the topic of free will. Which position do you find most convincing and why? Which do you consider to be its most challenging rival? Why is your position preferable to that other view?

## Suggestions for Further Reading

Free will has for a long time been one of the liveliest areas of research in contemporary metaphysics. There has been a lot written and there are lots of connections between this topic and related ones in the philosophy of mind and action, the philosophy of causation, and of course, related topics in ethics concerning moral responsibility. There are several nice books containing introductions to the topic, especially *Free Will: An Introduction* by Helen Beebe, *A Contemporary Introduction to Free Will* by Robert Kane, *Free Will: A Contemporary Introduction* by Michael McKenna and Derk Pereboom, and *Free Will: A Very Short Introduction* by Thomas Pink. P.F. Strawson's "Freedom and Resentment" is a classic and influential paper examining the deep connection between our judgment of others' actions as free and morally-loaded psychological attitudes (reactive attitudes) like resentment and guilt. The book *Four Views on Free Will* by John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas presents a defense of four positions on the topic: compatibilism, libertarianism, skepticism about free will, and revisionism (Vargas's view that our commonsense conception of free will needs to be substantially revised, but not abandoned) side by side. Van Inwagen presented his Consequence Argument in *An Essay on Free Will*. John Martin Fischer's reasons-responsive compatibilism is developed further in joint work with Mark Ravizza (1998). Carolina Sartorio develops a compatibilist account that emphasizes finding the right causes for actions in her book *Causation and Free Will*. In his "A Coherent, Naturalistic, and Plausible Formulation of Free Will" and *Free Will as an Open Scientific Problem*, Mark Balaguer defends libertarianism as the best answer to the scientific question about the causes of neural events, defending an event causal libertarianism distinct from Kane's. Derk Pereboom's recent *Wrongdoing and Moral Emotions* develops his account of how we can best understand our treatment of people who do bad things in the context of a hard incompatibilist position.

## Notes

- 1 One can find this distinction in the first chapter of Kane's *A Contemporary Introduction to Free Will*.
- 2 An earlier version appears in Ginet (1966).
- 3 Van Inwagen (2000b) considers a counterexample to  $\beta$  by Thomas McKay and David Johnson (1996). He agrees they have discovered a counterexample, and makes a slight modification to the inference rule that avoids it. Since the motivation for the change is rather technical and still lets the argument that follows proceed, we will stick with his original statement of the inference rule here.
- 4 This view has no relationship with the political libertarian position (the view in favor of a limited government) except for sharing a common name.
- 5 For an interesting discussion (and defense) of this possibility, see Nancy Cartwright's excellent book *The Dappled World*.

- 6 For a clear and comprehensive discussion of this debate, see David Z Albert's *Quantum Mechanics and Experience*.
- 7 Macroscopic objects are those large enough to be seen with the naked eye.
- 8 In Chapter 10, we discussed a couple of salient philosophical senses of the word 'possible': logically possible, nomologically possible. Lewis seems to be pressing the point that your having breakfast this morning was logically possible.
- 9 Though, van Inwagen himself has argued that Lewis's point does not require Humeanism in particular, just some other view about the relations between actions we undertake and the laws. See van Inwagen (2017).
- 10 See, for example, Slote (1982).
- 11 See McKenna and Pereboom (2014) and Timpe (2016) for a discussion of leeway vs. source freedom.
- 12 Frankfurt actually offered the case as a counterexample to what he called the Principle of Alternative Possibilities. This is the principle that one is only *morally responsible* for an act if one has the ability to do otherwise. So, this is strictly speaking not a principle about freedom but a principle about moral responsibility. However, there is a tight connection between the two phenomena. Many think that one can only be morally responsible for an action if it is free.
- 13 According to a 2020 survey of philosophers, 59.16% are compatibilists. 18.83% are libertarians, and 11.21% believe that no one has free will.
- 14 This isn't to say that the view is unproblematic or easy to understand. Indeed, even in Descartes's time (the seventeenth century) philosophers expressed puzzlement about how a nonphysical mental substance could affect anything in the physical world. See for example the correspondence between Descartes and Princess Elisabeth of Bohemia.
- 15 Although Kane seems influenced by cases like this from Sartre, it is worth noting that Sartre himself was not an event causal libertarian. Sartre's position is what one might call 'non-causalism.' Free actions are those that are not caused at all.
- 16 The butterfly effect is a phenomenon commonly used to illustrate chaotic behavior.

# Glossary of Terms

**A posteriori method:** an empirical way of knowing a fact or proposition, one that involves observation or sensory experience

**A priori method:** a way of knowing a fact or proposition that does not involve observation or sensory experience

**Abstract:** a classification of entities, examples include properties and mathematical objects

**Abstraction:** 1. A psychological process of considering an object while ignoring some of its features; for example, ignoring all other features of a table to just consider its size; 2. The metaphysical relationship of one entity being an abstraction from another, an entity just like the latter except lacking some of its features

**Abundant ontology:** an ontology that posits a relatively large number of types of entities

**Abundant theory of universals (or properties):** a version of realism about universals (or properties) according to which there exists a relatively large number of distinct universals (or properties); in one extreme case, a universal (or property) exists corresponding to any term that is applied to a multiplicity of entities

**Actualism:** the view that everything that exists actually exists

**A-features:** tensed features, features that change as time passes

**Agent causal libertarianism:** the view that human agents are sometimes causes and that is how their actions can be free

**Ameliorative project:** a metaphysical project that starts from the questions of what is the point of having a certain concept and what concept would do this work best

**Analytic:** truth in virtue of logic or meaning of the terms involved

**Analytic account:** an account of what we mean when we make some claim

**Argument:** a series of statements in which someone is presenting reasons in defense of some claim

**A-series:** a temporal ordering of events as being past, present, and future

**Asymmetry:** a relation *R* is asymmetric iff for all *x* and *y*, if *x* bears *R* to *y*, then *y* does not bear *R* to *x*

**Atheism:** the thesis that God does not exist

**A-theory of time:** the view that tensed facts are irreducible to and fail to be grounded in tenseless facts

**B-features:** tenseless but temporal features that do not change as time passes

**Biological determinism:** the view that the biological features of people of different sexes or genders determines significant facts about their psychologies and behaviors

**Biological realism about Xs:** the view that Xs are biological kinds

**Block universe view:** the combination of eternalism and the B-theory of time

**Bound variable:** a variable that is within the scope of some quantifier

**Brutal Composition:** the view that there is no true, interesting, and finite answer to the Special Composition Question

**B-series:** a temporal ordering of events as being earlier than, later than, or simultaneous with each other

**B-theory of time:** the view that tensed facts are reducible to or grounded in tenseless facts

**Category mistake:** a mistake that involves placing some entity in the wrong ontological category

**Class (or set) nominalism:** the view that properties are to be identified with the classes (or sets) of objects that instantiate them

**Cluster concept view:** a view about certain categories that although there is no set of necessary and sufficient conditions that all things that fall under the category must possess, the things that fall under that category do bear a family resemblance to one another, in possessing enough of a set of features

**Compatibilism:** the view that free will is compatible with determinism

**Conceptualism:** the view that universals exist, however they are entities that depend on our mind's grasp of them

**Conclusion:** the part of an argument that is being argued for, for which reasons are being offered

**Concrete:** a classification of entities that are not abstract; examples include material objects like tables, planets, and rocks

**Conferralism:** the view that a kind (or feature) *F* is socially constructed when its base feature *B* has social significance in a context



**Contingent:** neither necessary nor impossible

**Contingent a priori:** truths that are neither necessary nor impossible, and yet are discoverable merely by reflection on the meaning of the terms or concepts involved in them

**Contradiction:** any sentence or statement of the form  $P$  and not- $P$

**Conventionalism:** a view that seeks to reduce modal claims to claims about what follows or does not follow given the conventions of our language

**Counterexample:** an example that shows an argument is invalid, by providing a way in which the premises of the argument could be true, while the conclusion is false; or an example that shows a universal statement is false, by providing an instance in which it is false

**Counterfactual:** a conditional asserting what would have been the case had things gone differently than how we suppose they actually go

**Counterfactual theory of causation:** a theory that reduces facts about causation to facts about what would have happened in various counterfactual circumstances

**Counterpart:** a counterpart of one entity  $x$  is an entity that bears some salient similarity and causal relations to  $x$

**De dicto modality:** the topic of the status of propositions (or dictums) as possible, impossible, necessary, or contingent

**De re modality:** the topic of whether individual entities have properties that are essential to them or merely contingent

**Deductively invalid:** an argument is deductively invalid when it is possible for the premises of that argument to all be true while its conclusion is false

**Deductively valid:** an argument is deductively valid when there is no way for the premises to all be true while the conclusion is false. The premises of the argument logically imply its conclusion

**Descriptive metaphysics:** a metaphysical project of describing the actual structure of human thought about the world

**Determinism:** the view that the laws are such that, given any complete state of the universe at a time and facts about what the laws are, it is possible in principle to deduce what the complete state of the universe will be at any later time

**Diachronic identity:** identity over time

**Dilemma:** a choice between two options, each of which yields unattractive consequences

**Domain of quantification:** the set of objects over which the quantifiers range in a given context, the set of possible values the variable can take

**Efficient cause:** what brings an object or event into being

**Eleatic principle:** the principle that the entities that are real are those that are causally potent

**Eliminativism about Xs:** the view that Xs are not real

**Empiricism:** the view that our knowledge about the world comes entirely from experience

**Endurantism:** the view that what persistence amounts to is strict numerical identity over time

**Enthymeme:** an argument that is incomplete as stated and invalid, although it is easy to supply the missing premises that the argument would need to make it valid. In the case of an enthymeme, the author left out the missing premises for fear of boring the reader or insulting their intelligence

**Entrenchment:** a social role is entrenched when it tracks causal structures that are significant and so may support predictions and explanations

**Epiphenomenon:** an entity that lacks causal efficacy

**Epistemic possibility:** something that is compatible with everything that one knows

**Epistemic:** related to knowledge and truth

**Epistemicism:** the view that vagueness is a form of ignorance; it is not a matter of indeterminacy in the world or in what our words or concepts apply to, but our ignorance about what our words or concepts apply to

**Epistemology:** the theory of knowledge and justification

**Ersatz modal realism:** the view that there are possible worlds (worlds that play a similar role to the concrete worlds of the modal realist), but that these are not additional universes in the same sense as our universe

**Essence:** a set of features that make something the kind of thing it is, or the particular object that it is

**Essential property:** a property that holds of an entity with necessity and makes it the kind of entity that it is

**Essentialism:** the view that objects themselves, independently of any way we may categorize them, have certain properties necessarily

**Eternalism:** the view that past, present, and future objects and events are equally real

**Exdurantism (the stage theory):** the view that familiar, material objects are identical to temporary stages

**Existential quantifier:**  $\exists$ , a symbol of predicate logic. When combined with a variable, it can be used to represent a statement to the effect that something exists that is a certain way

**Existentialism:** the view that it is the kind of things we do that determines our essences, the kind of people we are; we do not possess innate essences that determine who we are and what we will do

**External question:** a question attempted to be interpreted from outside a particular linguistic framework

**External statement:** one attempted to be interpreted from outside a particular linguistic framework

**External time:** time itself

**Feminism:** a commitment to eradicating the ideology of domination that permeates Western culture on various levels – sex, race, and class, to name a few – and a commitment to reorganizing society so that the self-development of people can take precedence over imperialism, economic expansion, and material desires

**Fictionalism:** the view that truths in a certain domain should be understood as truths about a fiction

**Final (or teleological) cause:** the purpose or goal for which an object exists or why it is the way it is at a given time

**Forms:** the universals that constitute the fundamental entities of Plato's ontology

**Four dimensionalism:** the doctrine of temporal parts; the view that in addition to spatial parts, objects have temporal parts

**Framework (Carnapian):** a linguistic system including rules of grammar and meaning

**Frankfurt case:** a case in which intuitively one acts freely and so is morally responsible for an action, and yet one did not have the ability to do otherwise

**Genotype:** a set of genetic features

**Global Supervenience:** the A-properties globally supervene on the B-properties just in case worlds that are indiscernible with respect to the instantiation of B-properties ("B-indiscernible," for short) are also indiscernible with respect to the instantiation of A-properties (A-indiscernible)

**Ground:** a relation between facts or entities that makes for ontological dependence

**Growing block theory:** the view that past and present objects and events are real, but future ones are not

**Hard determinism:** the view that free will is incompatible with determinism, and so, since determinism is true, we lack free will

**Hard incompatibilism:** the view that free will is incompatible with both determinism and indeterminism and so we are not free

**Homeostatic mechanism:** a mechanism that ensures stability over time in such a way that either the presence of some properties tend to favor the presence of others, or there are underlying mechanisms or processes which tend to maintain the presence of the properties, or both

**Homeostatic property cluster account of natural kinds:** the view according to which members of a natural kind need not share a common set of intrinsic properties; rather, for there to be a natural kind, it is sufficient that there exist families of properties that are stably clustered in nature, where the stability of these clusters is due to the presence of a homeostatic mechanism

**Hylomorphism:** the Aristotelian view that substances are complex objects made of both matter (*hylē*) and form (*morphē*)

**Idealism:** the view that minds or mental entities (ideas) alone are fundamental

**Idealization:** a false assumption introduced into a theory in order to make it simpler to use

**Identity of Indiscernibles:** a principle stating that necessarily, if a and b are qualitative duplicates, then they are identical

**Identity theory of mind:** the view that mental properties are numerically identical with a certain type of physical properties, neurophysiological properties

**Immanent:** an entity that is located in space and time, where and when it is instantiated

**Incompatibilism:** the view that free will is incompatible with determinism

**Indispensability argument:** an argument for realism (Platonism) about mathematical entities from the premises that (1) we should be committed to the entities that are indispensable to our best scientific theories, and (2) the claim that mathematical entities are indispensable to our best scientific theories

**Inertial reference frames:** systems in which objects are moving at constant velocity

**Instantiation:** the relation between a property and an entity that has that property

**Interactive kind:** a kind that is the result of and in turn causes certain intentional actions

**Interactive substance dualism:** the view that minds and bodies are distinct kinds of substances that causally interact with one another

**Internal question:** a question asked and evaluated from within a specific linguistic framework

**Internal statement:** one that is stated and interpreted from within a linguistic framework

**Intersectionality:** the fact that the same person may be a member of several different social categories, so that they may be subordinated along several different dimensions, or subordinated in some respects, but privileged in others

**Intersubjectivity:** allowing for agreement by different people or subjects

**Intrinsic properties:** properties objects have just in virtue of how they are in themselves, not how they are in relation to other things

**Irreflexivity:** a relation *R* is irreflexive iff for any *x* it is false that *x* bears *R* to itself

**Leeway or alternative possibilities conception of freedom:** a conception of freedom according to which it requires the ability to do otherwise

**Leibniz's law (or the Indiscernibility of Identicals):** the principle that necessarily, if *a* and *b* are identical, then they must share all of the same properties

**Libertarianism:** the view that free will is incompatible with determinism and so, since we have free will, determinism is false

**Linguistic ersatzism:** a form of ersatz modal realism that takes possible worlds to be sentences or other linguistic entities

**Linguistic vagueness:** vagueness that is the result of semantic indecision; there not being facts to determine precisely in all cases what our terms apply to

**Local supervenience:** the *A*-properties locally supervene on the *B*-properties just in case individuals that are indiscernible with respect to the instantiation of *B*-properties ("B-indiscernible," for short) are also indiscernible with respect to the instantiation of *A*-properties (A-indiscernible)

**Logical connectives:** symbols used to build complex propositions out of simpler ones

**Logical positivism:** a movement in philosophy originating in Austria and Germany in the 1920s; a movement critical of metaphysics, arguing all knowledge must originate from sense experience or logic

**Logical possibility:** what does not entail any contradiction

**Logicism:** the view that mathematics is reducible to logic

**Major conclusion:** the final conclusion of an argument

**Manifest image:** the picture of the world we, as humans, work with in our ordinary lives using the concepts we have acquired and developed in order to explain our perceptual experiences, introspective sense of ourselves, and daily interactions

**Materialism:** a historical predecessor to contemporary physicalism, the view that matter alone is fundamental

**Maximal property:** a property *F* is maximal if large parts of an *F* are not themselves *F*s

**Mereological atom (or simple):** an object lacking proper parts

**Mereological nihilism:** the view that there are no mereologically complex objects, only simples

**Mereological sum (or fusion):** the mereological sum (or fusion) of some objects  $x_1, x_2, \dots, x_n$  is the object that contains  $x_1, x_2, \dots, x_n$  as parts

**Mereological universalism:** the view that any material objects, so long as they do not overlap spatially, compose some further object

**Meta-ontology:** the study of what one is doing, or what one should be doing, when one is engaged in an ontological debate

**Metaphysical explanation:** an explanation of a fact that says what there is in the world that accounts for that fact's being the case

**Metaphysical (or ontic) vagueness:** vagueness that results from how the world is objectively, not how we think or talk about it

**Mind-body dualism:** the view that there are two kinds of substances, minds (mental substances) and bodies (material substances)

**Minor conclusion:** a statement that is argued for on the way to arguing for an argument's major conclusion

**Mismatch objection:** the objection to biological theories of race that there appears to be a mismatch between our racial categories and the human continental populations

**Modal claim:** a claim that expresses a proposition about what is possible, impossible, necessary, or contingent

**Modal independence:** an entity is modally independent just in case its existence does not necessitate the existence of another entity

**Modal logic:** the branch of logic that deals with modal claims

**Modal properties:** properties having to do with what is possible, impossible, necessary, or contingent

**Modal realism:** the view that in addition to the actual world, there exist other alternative universes, possible worlds, just as real as our own; and that it is in virtue of the nature of these universes that our modal claims are true or false

**Model:** a theoretical structure involving a basic set of representational devices accounting for a set of data

**Modus ponens:** the logical form: If A, then B; A; Therefore, B

**Moving spotlight view:** the combination of eternalism and the A-theory of time

**Multiple realization argument:** an argument against type physicalism stating that because the same mental property can be and often is realized by distinct physical properties, mental properties cannot be identical to physical properties

**Natural kind:** a classification of objects that corresponds to objective joints in nature

**Natural properties:** a term introduced by David Lewis to distinguish those classes of objects in which each member is exactly and objectively similar to each other in some unique way

**Naturalism:** the view that it is within science itself that reality is to be identified and described

**Necessary a posteriori:** truths that are necessary, yet known on the basis of empirical observation

**Nominalism:** 1. The view that there are no such things as abstract entities; 2. The view that there are no such things as universals; 3. The view that there are no such things as mathematical entities

**Nomological possibility or necessity:** possibility or necessity according to the laws of nature

**Nonreductive physicalism:** the view that although physical properties are more fundamental than mental properties, mental properties are numerically distinct from physical properties

**Numbered premise form:** a way of stating arguments so that each premise as well as the conclusion are given a number and presented each on their own line

**Numerical identity (or identity in the strict sense):** oneness, the sense of 'a is identical to b' meaning that a and b are the same object, that they are one

**Objection from Coextension:** an argument against class nominalism that there are more properties than those that may be recognized by the class nominalist, since two predicates may have the same extension and yet refer to distinct properties

**Ockham's Razor:** the principle that one should not multiply one's ontological commitments beyond necessity

**One Over Many:** an argument for realism about universals that starts from a premise about some similarities between a group of objects and concludes that there is a universal (a one) that runs through these individual objects (the many)

**Ontological commitments:** the types of entities one ought to believe in, given the sentences one accepts

**Ontology:** 1. The study of what there is; 2. A particular theory about the types of entities there are

**Openness of the future:** the view that there are not any determinate facts about the future

**Origins essentialism:** the view that the origins of material objects and organisms are essential to them

**Ostrich nominalism:** a version of nominalism that denies the existence of properties and refuses to answer the question of what it is in virtue of which objects are similar or appear to have certain features

**Particular:** a type of entity that is not capable of multiple instantiation

**Perdurantism (the worm view):** the view that material objects persist by having temporal parts at different times

**Personal time:** elapsed time as measured by the normal behavior of physical objects

**Phenomenalism:** the view that mental properties are more fundamental than physical properties

**Phenotype:** a set of observable features

**Physicalism:** the view that physics alone can provide a complete description of the fundamental nature of our world

**Platonism:** 1. The view that there are such things as the Platonic Forms; 2. The view that there are such things as abstract, mathematical entities

**Positional view of gender:** the view that genders are real, but having a gender isn't having a set of intrinsic features, but rather occupying a kind of social position

**Possibilism:** the view that there are at least some entities that are merely possible

**Possible worlds analysis of modality:** an analysis of claims about possibility and necessity in terms of what is true at various possible worlds (including the actual world)

**Predicate nominalism:** a view denying the existence of properties; predicates may be satisfied or not satisfied by objects, but there need be no property that exists to explain this fact

**Preestablished harmony:** mental and physical properties are correlated not because they ontologically depend on one another, but because the correlation is set up that way by God

**Premise:** a statement offered as part of an argument as a reason for accepting a certain claim

**Presentism:** the view that only present objects and events are real

**Primitive concept:** a concept is primitive if it does not have an analysis in terms of other concepts

**Primitivist theory of causation:** a theory according to which causal facts are not reducible to any noncausal facts, including facts about regularities, laws, counterfactuals, or probabilities

**Principle of charity:** a convention of philosophical debate to, where reasonable, try to interpret one's opponent's claims as true and their arguments as valid



**Principle of Naturalistic Closure:** the principle that any metaphysical claim to be taken seriously at a time should be motivated by the service it would perform in showing how two or more scientific hypotheses, at least one of which is drawn from fundamental physics, jointly explain more than what is explained by the hypotheses taken separately

**Priority monism:** the view that the one ultimate whole is more fundamental than its parts

**Priority pluralism:** the view that parts are more fundamental than wholes

**Problem of essentialism:** a problem for accounts of gender of reinforcing essentialism, so that those who do not meet the conditions of the account are lesser or inferior members of that gender

**Problem of temporary intrinsics:** a problem for endurantism that it cannot account for change in an object's intrinsic properties

**Problem of the Many:** a philosophical problem about the existence and identity of material objects that arises from the fact that they seem not to have well-defined boundaries; if there are several precisely defined objects with determinate boundaries in the vicinity of a material object, this raises the question of which if any that object is identical to

**Projectible:** a kind is projectible if you can reliably predict what things of that kind will do in future circumstances

**Proper part:** x is a proper part of y just in case x is a part of y and x is not identical to y

**Property dualism:** the view that mental properties and physical properties are numerically distinct and equally fundamental

**Protocol statement:** a statement that may be directly verified by sense experience

**Qualitative identity:** the sense of 'a is identical to b' meaning that a and b share qualities (the same color, the same size, and so on)

**Qualitative parsimony:** the degree to which an ontology includes fewer types of things

**Quantitative parsimony:** the degree to which an ontology includes a fewer number of things

**Racialism:** the view that there are heritable characteristics which divide human beings into a small set of races, in such a way that all the members of these races share certain culturally significant traits and tendencies with each other that they do not share with members of any other race

**Realism about universals:** the view that universals exist and they are mind-independent entities

**Realization:** a relation between two properties A and B such that B is the property that plays the causal/functional role characteristic of property A

**Reductio ad absurdum:** the method of proving a claim by showing that the negation of that claim entails a contradiction

**Reference class problem:** the problem that the probability we assign to an event's occurring seems to depend on our way of conceptualizing it (placing it against a reference class) on a given occasion; since the reference class we use can depend on the context, this makes it difficult to say what is the probability of an event's occurring

**Reflexivity:** a relation  $r$  is reflexive iff for all  $x$ ,  $x$  bears  $R$  to itself

**Regimentation:** the procedure of representing statements in symbolic logic to make it as clear as possible what follows from these statements

**Regularity theory of causation:** a theory of causation that explains causal relations in terms of the regular occurrence of patterns of events

**Relative fundamentality:** one entity or fact's being more, less, or equally fundamental than some other entity or fact

**Relativity objection:** the objection to biological theories of race that different cultures have different racial categories, and yet if races are natural kinds, then they must not vary with culture

**Revisionary metaphysics:** a metaphysical project of improving the structure of human thought about the world

**Scientific image:** the picture of the world science provides

**Scope (of a quantifier):** the part of a sentence containing the variables the quantifier is binding. In symbolic logic, the scope of a quantifier is either the part of the sentence immediately after the quantifier phrase (in a simple sentence like ' $\exists xFx$ ') or the part of the sentence contained in the parentheses that immediately follow the quantifier phrase. For example, in ' $\exists x(Fx \wedge Gx) \wedge Hx$ ,' the  $x$ s in ' $Fx$ ' and ' $Gx$ ' are in the scope of the quantifier. The  $x$  in ' $Hx$ ' is not

**Self-forming actions:** important actions in the life of a person that decide the kind of person they will be

**Semantic ascent:** when, in order to address a question, a philosopher "ascends to the semantic plane," addressing first a question about the meaning of key terms in the original question

**Semantic theory:** an account of a proposition's or a set of propositions' meaning and truth conditions

**Sentential operator:** an operator acting on an entire sentence or proposition

**Social constructionism about Xs:** the view that Xs are social kinds

**Social kinds:** kinds that depend for their existence on human behavior or interactions

**Social significance:** a feature  $B$  has social significance in a context when another feature  $F$  is conferred upon people taken to have  $B$

**Soft determinism:** the view that determinism is true and it is compatible with the existence of free will

**Sortal essentialism:** the view that it is essential to objects what kind of things they are

**Sound:** an argument is sound just in case it has all true premises and is deductively valid

**Sourcehood conception of freedom:** a conception of freedom according to which it requires an agent to be the proper causal source of their actions

**Sparse ontology:** an ontology that posits a relatively small number of types of entities

**Sparse theory of universals (or properties):** a version of realism about universals (or properties) according to which there exists a relatively small number of universals (or properties); in one extreme case, there are only universals corresponding to the types recognized by our most fundamental physical theories

**Special Composition Question:** the question for any  $x$ s, when it is the case that there is a  $y$  such that the  $x$ s compose  $y$

**Strong Supervenience:** A strongly supervenes on B just in case, necessarily, for each  $x$  and each property  $F$  in A, if  $x$  has  $F$ , then there is a property  $G$  in B such that  $x$  has  $G$ , and *necessarily* if any  $y$  has  $G$ , it has  $F$

**Supervenience:** a relation between two sets of properties A and B such that the instantiation of the A-properties completely fix the instantiation of the B-properties

**Surface freedom:** being able to act in such a way that one's desires are satisfied

**Synchronic identity:** identity at a time

**Synthetic:** empirically verifiable

**Theism:** the thesis that God exists

**Thought experiment:** a fictional case used to draw out consequences of use to the building of a scientific or philosophical theory

**Three dimensionalism:** the view that although objects may have spatial parts, they never have temporal parts

**Transcendent:** a transcendent entity is one that is not located in space or time

**Transitivity:** a relation  $R$  is transitive iff for all  $x$ ,  $y$ , and  $z$ , if  $x$  bears  $R$  to  $y$ , and  $y$  bears  $R$  to  $z$ , then  $x$  bears  $R$  to  $z$

**Trope:** an abstract particular, e.g. the shape of the Empire State Building

**Trope theory (or the theory of abstract particulars):** the theory that properties are tropes, or abstract particulars

**Truthmaker maximalism:** the view that all truths have truthmakers

**Truthmaker theory:** the theory that truths have truth makers, some entities or sets of entities that make them true

**Two Object View:** the view that material objects are numerically distinct from the matter of which they are made

**Ultimate freedom:** having the ability to not only satisfy one's desires but being the ultimate source of these desires

**Universal:** a type of entity that is repeatable, that may be instantiated at multiple locations at once by distinct objects

**Universal quantifier:**  $\forall$ , a symbol of predicate logic. When combined with a variable, it can be used to represent a statement to the effect that everything is a certain way

**Use/mention distinction:** a distinction between two ways in which a word or phrase may appear in a sentence. A sentence may use the linguistic item so that it plays its typical semantic role (naming some object if it is a name, modifying some object if it is an adjective, and so on). Or, a sentence may mention the linguistic item, using it to refer to itself. In cases where a linguistic item is being mentioned, a philosophical convention is to place the relevant word or phrase in single quotes

**Variables:** symbols like  $x$ ,  $y$ ,  $z$ , etc. used to stand in for other things in a sentence, called the values of the variable

**Verificationist theory of meaning:** the meaning of a statement is given by its verification conditions

**Verificationist theory of truth:** a sentence is only capable of truth or falsity if it is able to be verified or falsified

**Weak Supervenience:** A weakly supervenes on B if and only if necessarily, for any  $x$  and  $y$ , if  $x$  and  $y$  share all properties in B, then  $x$  and  $y$  share all properties in A

**Wholly present:** to be wholly present at a time is to have all of one's parts exist at that time

**Worldline:** the path of any object through spacetime

# Works Cited

- Akiba, Ken. 2004. Vagueness in the World. *Noûs*. 38(3): 407-429.
- Albert, David Z. 1992. *Quantum Mechanics and Experience*. Cambridge: Harvard University Press.
- Alcoff, Linda Martín. 2006. *Visible Identities: Race, Gender, and the Self*. New York: Oxford University Press.
- Alcoff, Linda Martín. 2015. *The Future of Whiteness*. Cambridge: Polity.
- Alston, William. 1958. Ontological Commitments. *Philosophical Studies*. 9(1-2): 8-17.
- Andreasen, Robin. 1998. A New Perspective on the Race Debate. *British Journal for the Philosophy of Science*. 49(2): 199-225.
- Annas, Julia. 1981. *An Introduction to Plato's Republic*. Oxford: Clarendon Press.
- Anscombe, G.E.M. 1957. *Intention*. Oxford: Basil Blackwell.
- Appiah, Kwame Anthony. 1992. *In My Father's House: Africa in the Philosophy of Culture*. New York: Oxford University Press.
- Aristotle. 1975. *Categories and De Interpretatione*. (Translator: J.M. Ackrill.) Oxford: Oxford University Press.
- Armstrong, D.M. 1978a. *Nominalism and Realism: Universals and Scientific Realism Volume 1*. Cambridge: Cambridge University Press.
- Armstrong, D.M. 1978b. *A Theory of Universals: Universals and Scientific Realism Volume 2*. Cambridge: Cambridge University Press.
- Armstrong, D.M. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.
- Armstrong, D.M. 1989a. *Universals: An Opinionated Introduction*. Denver: Westview Press.
- Armstrong, D.M. 1989b. *A Combinatorial Theory of Possibility*. Cambridge: Cambridge University Press.
- Armstrong, D.M. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Armstrong, D.M. 2004. *Truths and Truthmakers*. Cambridge: Cambridge University Press.
- Ásta. 2018. *Categories We Live By*. New York: Oxford.
- Audi, Paul. 2012. Grounding: Toward a Theory of the In-Virtue-Of Relation. *Journal of Philosophy*. 109(12): 685-711.
- Ayer, A.J. 1936. *Language, Truth, and Logic*. London: Gollancz.
- Ayer, A.J., ed. 1959. *Logical Positivism*. New York: The Free Press.
- Ayer, A.J. 1969. Freedom and Necessity. In *Philosophical Essays*. London: Macmillan.
- Bach, Theodore. 2012. Gender is a Natural Kind with a Historical Essence. *Ethics*. 122(2): 231-272.
- Baker, Alan. 2005. Are There Genuine Mathematical Explanations of Physical Phenomena? *Mind*. 114(454): 223-238.
- Baker, Alan. 2022. Simplicity. In *The Stanford Encyclopedia of Philosophy*. (Winter 2016 Edition.) E. Zalta, ed. URL <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=simplicity>

- Balaguer, Mark. 2004. A Coherent, Naturalistic, and Plausible Formulation of Libertarian Free Will. *Noûs*. 38(3): 379-406.
- Balaguer, Mark. 2010. *Free Will as an Open Scientific Problem*. Cambridge: MIT Press.
- Balaguer, Mark. 2021. *Metaphysics, Sophistry, and Illusion: Toward a Widespread Non-Factualism*. Oxford: Oxford University Press.
- Barnes, Elizabeth. 2010. Ontic Vagueness: A Guide for the Perplexed. *Noûs*. 44(4): 601-627.
- Barnes, Elizabeth. 2012. Emergence and Fundamentality. *Mind*. 121(484): 873-901.
- Barnes, Elizabeth. 2016. *The Minority Body*. Oxford: Oxford University Press.
- Barnes, Elizabeth. 2018. Symmetric Dependence. In Bliss and Priest 2018.
- Barnes, Elizabeth and J. Robert G. Williams. 2011. A Theory of Metaphysical Indeterminacy. *Oxford Studies in Metaphysics*. 6: 103-148.
- Beauvoir, Simone de. 1949. *The Second Sex*. Paris: Librairie Gallimard.
- Beebee, Helen. 2013. *Free Will: An Introduction*. London: Palgrave Macmillan.
- Benacerraf, Paul. 1973. Mathematical Truth. *Journal of Philosophy*. 70(19): 661-679.
- Benacerraf, Paul and Hilary Putnam. 1964. *Philosophy of Mathematics: Selected Readings*. Cambridge: Cambridge University Press.
- Bennett, Karen. 2004. Spatio-Temporal Coincidence and the Grounding Problem. *Philosophical Studies*. 118(3): 339-371.
- Bennett, Karen. 2017. *Making Things Up*. Oxford: Oxford University Press.
- Bergmann, Merrie, James Moor, and Jack Nelson. 2008. *The Logic Book*. New York: McGraw-Hill.
- Bettcher, Talia Mae. 2009. Trans Identities and First-Person Authority. In *You've Changed: Sex Reassignment and Personal Identity*. L. Shrage, ed. Oxford: Oxford University Press, 98-120.
- Bettcher, Talia Mae. 2011. Trans Women and the Meaning of "Woman." *The American Philosophical Association Newsletter on Philosophy and Lesbian, Gay, Bisexual, and Transgender Issues*. 10(2): 2-5.
- Bigelow, John. 1996. Presentism and Properties. *Philosophical Perspectives*. 10: 35-52.
- Black, Max. 1952. The Identity of Indiscernibles. *Mind*. 61(242): 153-164.
- Bliss, Ricki. 2018. Grounding and Reflexivity. In Bliss and Priest 2018.
- Bliss, Ricki and Graham Priest, eds. 2018. *Reality and Its Structure: Essays in Fundamentality*. Oxford: Oxford University Press.
- Block, Ned and Robert Stalnaker. 1999. Conceptual Analysis, Dualism, and the Explanatory Gap. *Philosophical Review*. 108(1): 1-46.
- Boolos, George. 1984. To Be is to Be a Value of a Variable (or to Be Some Values of Some Variables). *Journal of Philosophy*. 81(8): 430-449.
- Borges, Jorge Luis. 1964. The Analytical Language of John Wilkins. In *Other Inquisitions 1937-1952*. Austin: University of Texas Press.
- Boyd, Richard. 1999. Homeostasis, Species, and Higher Taxa. In *Species: New Interdisciplinary Essays*. R.A. Wilson, ed. Cambridge: MIT Press, 141-185.
- Burgess, Alexis, Herman Cappelen, and David Plunkett. 2020. *Conceptual Engineering and Conceptual Ethics*. Oxford: Oxford University Press.
- Butler, Judith. 1990. *Gender Trouble*. London: Routledge.
- Calosi, Claudio. 2018. Quantum Monism: An Assessment. *Philosophical Studies*. 175(12): 3217-3236.
- Cameron, Ross. 2008. Turtles All the Way Down: Regress, Priority, and Fundamentality. *Philosophical Quarterly*. 58(230): 1-14.
- Campbell, Joseph Keim, Michael O'Rourke, and Matthew H. Slater, eds. 2011. *Carving Nature at Its Joints*. Cambridge: MIT Press.
- Carnap, Rudolf. 1932. The Elimination of Metaphysics Through the Logical Analysis of Language. *Erkenntnis*. 2: 60-81.
- Carnap, Rudolf. 1950. Empiricism, Semantics, and Ontology. *Revue Internationale de Philosophie*. 4(11): 20-40.
- Cartwright, Nancy. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
- Cartwright, Richard. 1968. Some Remarks on Essentialism. *Journal of Philosophy*. 65(2): 615-626.

- Chalmers, David. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Chalmers, David. 2012. *Constructing the World*. Oxford: Oxford University Press.
- Chisholm, Roderick. 1964. Human Freedom and the Self. Lindley Lecture. University of Kansas.
- Churchland, Paul M. 1985. *Matter and Consciousness*. Cambridge: MIT Press.
- Collins, J., N. Hall, and L.A. Paul, eds. 2004. *Causation and Counterfactuals*. Oxford: Oxford University Press.
- Colyvan, Mark. 2001. *The Indispensability of Mathematics*. Oxford: Oxford University Press.
- Correia, Fabrice and Benjamin Schnieder, eds. 2012. *Metaphysical Grounding*. Cambridge: Cambridge University Press.
- Cowling, Sam. 2013. Ideological Parsimony. *Synthese*. 190(17): 3889-3908.
- Cowling, Sam. 2017. *Abstract Entities*. London: Routledge.
- Craig, William Lane. 2000. *The Tensed Theory of Time: A Critical Examination*. Dordrecht: Springer.
- Craig, William Lane. 2000. *The Tensed Theory of Time*. Dordrecht: Kluwer.
- Craig, William Lane, ed. 2001. *Time and the Metaphysics of Relativity*. Dordrecht: Kluwer.
- Craver, Carl. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.
- Daly, Chris. 2012. Scepticism about Grounding. In Correia and Schnieder 2012.
- Daston, Lorraine and Peter Galison. 2007. *Objectivity*. Brooklyn: Zone Books.
- Davidson, Donald. 1967a. The Logical Form of Action Sentences. In *The Logic of Decision and Action*. N. Rescher, ed. Pittsburgh: University of Pittsburgh Press, 81-95.
- Davidson, Donald. 1967b. Causal Relations. *Journal of Philosophy*. 64(21): 691-703.
- Davidson, Donald. 1970. Events as Particulars. *Noûs*. 4(1): 25-32.
- Dembroff, Robin. 2018. Real Talk on the Metaphysics of Gender. *Philosophical Topics*. 46(2): 21-50.
- Dennett, Daniel C. 2003. *Freedom Evolves*. New York: Viking.
- Descartes, Rene. 1641. *Meditations on First Philosophy*.
- Descartes, Rene and Princess Elisabeth of Bohemia. 2007. *The Correspondence between Princess Elisabeth of Bohemia and Rene Descartes*. Chicago: University of Chicago Press.
- Díaz León, Esa. 2020. Descriptive vs. Ameliorative Projects. In Burgess, Cappelen, and Plunkett 2020.
- Dowe, Phil. 2000. *Physical Causation*. Cambridge: Cambridge University Press.
- Du Bois, W.E.B. 1897. The Conservation of the Races. In *W.E.B. Du Bois Speaks: Speeches and Address 1890-1919*. P.S. Foner, ed. New York: Pathfinders, 73-85.
- Dupre, John. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge: Harvard University Press.
- Effingham, Nikk. 2020. *Time Travel: Probability and Impossibility*. Oxford: Oxford University Press.
- Einstein, Albert. 1916/2013. *Relativity: The Special and General Theory*. New York: Empire Books.
- Epstein, Brian. 2015. *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. Oxford: Oxford University Press.
- Epstein, Brian. 2018. Social Ontology. *Stanford Online Encyclopedia of Philosophy*.
- Evans, Gareth. 1978. Can There Be Vague Objects? *Analysis*. 38(4): 208.
- Fausto-Sterling, Anne. 1993. The Five Sexes: Why Male and Female are Not Enough. *The Sciences*. 33: 20-24.
- Feigl, Herbert. 1958. The "Mental" and the "Physical." *Minnesota Studies in the Philosophy of Science*. 2. Minneapolis: University of Minnesota Press.
- Feldman, Richard. 1998. *Reason and Argument*. New York: Prentice Hall.
- Field, Hartry. 1980. *Science Without Numbers: A Defence of Nominalism*. Princeton: Princeton University Press.
- Fine, Kit. 1994. Essence and Modality. *Philosophical Perspectives*. 8: 1-16.

- Fine, Kit. 2001. The Question of Realism. *Philosophers' Imprint*. 1: 1-30.
- Fine, Kit. 2005. Tense and Reality. In *Modality and Tense: Philosophical Papers*. Oxford: Clarendon Press, 261-230.
- Fine, Kit. 2012. Guide to Ground. In Correia and Schnieder 2012: 37-80.
- Fischer, John Martin. 1994. *The Metaphysics of Free Will*. Oxford: Blackwell.
- Fischer, John Martin, Robert Kane, Derk Pereboom, and Manuel Vargas. 2007. *Four Views on Free Will*. Oxford: Blackwell.
- Fischer, John Martin and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Fodor, Jerry. 1974. Special Sciences. *Synthese*. 28(2): 97-115.
- Forrest, Peter. 1986. Ways Worlds Could Be. *Australasian Journal of Philosophy*. 64(1): 15-24.
- Frankfurt, Harry. 1969. Alternative Possibilities and Moral Responsibility. *Journal of Philosophy*. 66(3): 829-839.
- Frankfurt, Harry. 1971. Freedom of the Will and the Concept of a Person. *Journal of Philosophy*. 68: 5-20.
- French, Steven and Kerry McKenzie. 2012. Thinking Outside the Toolbox: Towards a More Productive Engagement Between Metaphysics and Philosophy of Physics. *European Journal of Analytic Philosophy*. 8(1): 42-59.
- Friedan, Betty. 1963. *The Feminist Mystique*. London: W.W. Norton and Company.
- Gendler, Tamar Szabó and John Hawthorne, eds. 2002. *Conceivability and Possibility*. Oxford: Oxford University Press.
- Gillett, Carl. 2002. The Dimensions of Realization: A Critique of the Standard View. *Analysis*. 62(4): 316-323.
- Gilmore, Cody and Matt Leonard. 2020. Composition and the Logic of Location: An Argument for Regionalism. *Mind*. 129(513): 159-178.
- Ginet, Carl. 1966. Might We Have No Choice? In *Freedom and Determinism*. K. Lehrer, ed. New York: Random House, 87-104.
- Glasgow, Joshua. 2009. *A Theory of Race*. London: Routledge.
- Glasgow, Joshua, Sally Haslanger, Chike Jeffers, and Quayshawn Spencer. 2019. *What is Race? Four Philosophical Views*. New York: Oxford.
- Goff, Philip. 2017. *Consciousness and Fundamental Reality*. Oxford: Oxford University Press.
- Hacking, Ian. 1999. *The Social Construction of What?* Cambridge: Harvard University Press.
- Hall, Ned and L.A. Paul. 2013. *Causation: A User's Guide*. Oxford: Oxford University Press.
- Hardegree, Gary. 1999. *Symbolic Logic: A First Course*. New York: McGraw-Hill.
- Harris, Sam. 2012. *Free Will*. New York: Free Press.
- Haslanger, Sally. 2003. Persistence Through Time. In *The Oxford Handbook of Metaphysics*. M. Loux and D. Zimmerman, eds. Oxford: Oxford University Press.
- Haslanger, Sally. 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.
- Haslanger, Sally and Roxanne Marie Kurtz. 2006. *Persistence: Contemporary Readings*. Cambridge: MIT Press.
- Hawley, Katherine. 2010. Temporal Parts. *The Stanford Encyclopedia of Philosophy* (Summer 2020) Edition. E. Zalta, ed. URL <https://plato.stanford.edu/archives/sum2020/entries/temporal-parts/>
- Hawley, Katherine and Alexander Bird. 2011. What Are Natural Kinds? *Philosophical Perspectives*. 25(1): 205-221.
- Hawthorne, John and Daniel Nolan. 2006. *What Would Teleological Causation Be?* In *Metaphysical Essays*. Oxford: Oxford University Press.
- Heidegger, Martin. 1993. What is Metaphysics? In *Martin Heidegger: Basic Writings*. D.F. Krell, ed. London: Routledge.
- Heil, John. 2012. *The Universe as We Find It*. Oxford: Oxford University Press.
- Heller, Mark. 1990. *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter*. Cambridge: Cambridge University Press.



- Hestevold, H. Scott. 1981. Conjoining. *Philosophy and Phenomenological Research*. 41(3): 371-385.
- Hill, Christopher S. 1991. *Sensations: A Defense of Type Materialism*. Cambridge: Cambridge University Press.
- Hinchliff, Mark. 2000. A Defense of Presentism in a Relativistic Setting. *Philosophy of Science*. 67(3): S575-S586.
- Hirsch, Eli. 2010. *Quantifier Variance and Realism: Essays in Metaontology*. Oxford: Oxford University Press.
- Hobbes, Thomas. 1651. *Leviathan*. New York: Norton.
- Hofweber, Thomas. 2009. Ambitious, yet Modest, Metaphysics. In *Metametaphysics*. D. Chalmers, Manley, and Wasserman 2009.
- Hofweber, Thomas and Marc Lange. 2017. Fine's Fragmentalist Interpretation of Special Relativity. *Noûs*. 51(4): 871-883.
- hooks, bell. 1982. *Ain't I a Woman: Black Women and Feminism*. Boston: South End Press.
- Horwich, Paul. 1975. On Some Alleged Paradoxes of Time Travel. *Journal of Philosophy*. 72(14): 32-444.
- Howell, Robert. 2009. Emergentism and Supervenience Physicalism. *Australasian Journal of Philosophy*. 87(1): 83-98.
- Hume, David. 1861. *An Enquiry Concerning Human Understanding*. London: Bebbington.
- Hylton, Peter. 2007. *Quine*. London: Routledge.
- Ismael, Jenann and Jonathan Schaffer. 2020. Quantum Holism: Nonseparability as Common Ground. *Synthese*. 197(10): 4131-4160.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
- Jeffers, Chike. 2019. Cultural Construction. In *What is Race? Four Philosophical Views*. Oxford: Oxford University Press.
- Jenkins, Katherine. 2016. Amelioration and Inclusion: Gender Identity and the Concept of Woman. *Ethics*. 126(2): 394-421.
- Kalderon, Mark, ed. 2005. *Fictionalism in Metaphysics*. Oxford: Oxford University Press.
- Kane, Robert. 1996. *The Significance of Free Will*. New York: Oxford University Press.
- Kane, Robert. 2005. *A Contemporary Introduction to Free Will*. Oxford: Oxford University Press.
- Keefe, Rosanna and Peter Smith. 1997. *Vagueness: A Reader*. Cambridge: MIT Press.
- Keller, Simon. 2004. Presentism and Truthmaking. *Oxford Studies in Metaphysics*. 1: 83-104.
- Keller, Simon and Michael Nelson. 2001. Presentists Should Believe in Time Travel. *Australasian Journal of Philosophy*. 79(3): 333-345.
- Kenny, Anthony. 1963. *Action, Emotion, and Will*. Oxford: Wiley-Blackwell.
- Khalidi, Muhammad Ali. 2015. Three Kinds of Social Kinds. *Philosophy and Phenomenological Research*. 90(1): 96-112.
- Kim, Jaegwon. 1976. Events as Property Exemplifications. In *Action Theory*. M. Brand and D. Walton, eds. Dordrecht: Reidel.
- Kim, Jaegwon. 1984. Concepts of Supervenience. *Philosophy and Phenomenological Research*. 45: 153-176.
- Kim, Jaegwon. 1996. *Philosophy of Mind*. Westview Press.
- Kim, Jaegwon. 1998. *Mind in a Physical World*. Cambridge: MIT Press.
- Kim, Jaegwon. 2003. The American Origins of Philosophical Naturalism. *Journal of Philosophical Research*. 28(Supplement): 83-98.
- Kim, Jaegwon. 2005. *Physicalism, Or Something Near Enough*. Princeton: Princeton University Press.
- Kleinschmidt, Shieva, ed. 2014. *Mereology and Location*. Oxford: Oxford University Press.

- Koslicki, Kathrin. 2008. *The Structure of Objects*. Oxford: Oxford University Press.
- Koslicki, Kathrin. 2015. The Coarse-Grainedness of Grounding. *Oxford Studies in Metaphysics*. 9: 306-44.
- Kripke, Saul. 1970/1980. *Naming and Necessity*. Cambridge: Harvard University Press.
- Kuhn, Thomas. 1977. Objectivity, Value Judgment, and Theory Choice. In *The Essential Tension*. Chicago: University of Chicago Press.
- Ladyman, James. 2007. Does Physics Answer Metaphysical Questions. *Royal Institute of Philosophy Supplement*. 61: 179-201.
- Ladyman, James. 2012. Science, Metaphysics, and Method. *Philosophical Studies*. 160(1): 31-51.
- Ladyman, James. 2017. An Apology for Naturalized Metaphysics. In *Metaphysics and the Philosophy of Science*. M. Slater and Z. Yudell, eds. New York: Oxford University Press.
- Ladyman, James and Don Ross. 2007. *Every Thing Must Go*. Oxford: Oxford University Press.
- Langton, Rae and David Lewis. 1998. Defining 'Intrinsic.' *Philosophy and Phenomenological Research*. 58(2): 333-345.
- Laplace, Pierre-Simon. 1951. *A Philosophical Essay on Probabilities*. New York: Dover.
- Leonard, Henry S. and Nelson Goodman. 1940. The Calculus of Individuals and Its Uses. *Journal of Symbolic Logic*. 5(2): 45-55.
- Le Poidevin, Robin and Murray MacBeath. 1993. *The Philosophy of Time*. Oxford: Oxford University Press.
- Lewis, David. 1969. *Convention: A Philosophical Study*. Newcastle: Basil Blackwell.
- Lewis, David. 1972. Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*. 50(3): 249-258.
- Lewis, David. 1973a. Counterfactuals and Comparative Similarity. *Journal of Philosophical Logic*. 2(4): 418-446.
- Lewis, David. 1973b. Causation. *Journal of Philosophy*. 70(17): 556-567. Reprinted with Postscripts in *Philosophical Papers: Volume II*, 1986. Oxford: Oxford University Press, 159-213.
- Lewis, David. 1976. The Paradoxes of Time Travel. *American Philosophical Quarterly*. 13(2): 125-152.
- Lewis, David. 1978. Truth in Fiction. *American Philosophical Quarterly*. 15(1): 37-46.
- Lewis, David. 1981. Are We Free to Break the Laws? *Theoria*. 47(3): 113-121.
- Lewis, David. 1983. New Work for a Theory of Universals. *Australasian Journal of Philosophy*. 61(4): 343-377.
- Lewis, David. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.
- Lewis, David. 1993. Many, But Almost One. In *Ontology, Causality and Mind: Essays on the Philosophy of D.M. Armstrong*. Cambridge: Cambridge University Press, 23-38.
- Lewis, David. 2000. Causation as Influence. *Journal of Philosophy*. 97(4): 182-197.
- Libet, B., C.A. Gleason, E.W. Wright, and D.K. Pearl. 1983. Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act. *Brain*. 106(3): 623-642.
- Loux, Michael, ed. 1979. *The Possible and the Actual*. Ithaca: Cornell University Press.
- Lowe, E.J. 2005. *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Oxford: Clarendon Press.
- Mackie, J.L. 1980. *The Cement of the Universe*. Oxford: Clarendon.
- Mackie, Penelope. 2006. *How Things Might Have Been: Individuals, Kinds, and Essential Properties*. Oxford: Oxford University Press.
- Maddy, Penelope. 1992. Indispensability and Practice. *Journal of Philosophy*. 89(6): 275-289.
- Maddy, Penelope. 2007. *Second Philosophy*. Oxford: Oxford University Press.
- Magnus, P.D. 2012. *Scientific Enquiry and Natural Kinds*. London: Palgrave Macmillan.
- Mallon, Ron. 2016. *The Construction of Human Kinds*. Oxford: Oxford University Press.

- Marcus, Ruth Barcan. 1967. Essentialism in Modal Logic. *Noûs*. 1(1): 91-96.
- Marcus, Ruth Barcan. 1990. A Backward Look at Quine's Animadversions on Modalities. In *Perspectives on Quine*. R. Barrett and R. Gibson, eds. Oxford: Blackwell, 230-243.
- Markosian, Ned. 1998. Brutal Composition. *Philosophical Studies*. 92(3): 211-249.
- Markosian, Ned. 2004. A Defense of Presentism. *Oxford Studies in Metaphysics*. 1: 47-82.
- Markosian, Ned. 2013. The Truth About the Past and the Future. *Around the Tree: Semantic and Metaphysical Issues Concerning Branching Time and the Open Future*. F. Correia and A. Iacona, eds. Dordrecht: Springer, 127-141.
- Markosian, Ned. 2014. A Spatial Approach to Mereology. *Mereology and Location*. S. Kleinschmidt, ed. Oxford: Oxford University Press.
- Maudlin, Tim. 2007. *The Metaphysics Within Physics*. Oxford: Oxford University Press.
- McDaniel, Kris. 2007. Extended Simples. *Philosophical Studies*. 133(1): 131-141.
- McDaniel, Kris. 2017. *The Fragmentation of Being*. Oxford: Oxford University Press.
- McKay, Thomas. 1999. *Reasons, Explanations, and Decisions: Guidelines for Critical Thinking*. London: Wadsworth.
- McKay, Thomas and David Johnson. 1996. A Reconsideration of an Argument against Compatibilism. *Philosophical Topics*. 24: 113-122.
- McKenna, Michael. 2008. A Hard-Line Response to Pereboom's Four-Case Argument. *Philosophy and Phenomenological Research*. 77: 142-159.
- McKenna, Michael and Derk Pereboom. 2014. *Free Will: A Contemporary Introduction*. London: Routledge.
- McTaggart, J.M.E. 1908. The Unreality of Time. *Mind*. 17(68): 457-474.
- Meiland, Jack. 1974. Two-Dimensional Passage Model of Time for Time Travel. *Philosophical Studies*. 26(3-4): 153-173.
- Mele, Alfred. 2006. *Free Will and Luck*. New York: Oxford University Press.
- Mellor, D.H. 1995. *The Facts of Causation*. London: Routledge.
- Mellor, D.H. and Alex Oliver, eds. 1997. *Properties*. Oxford: Oxford University Press.
- Melnyk, Andrew. 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*. New York: Cambridge University Press.
- Merricks, Trenton. 2001. *Objects and Persons*. Oxford: Oxford University Press.
- Mikkola, Mari. 2017. Feminist Perspectives on Sex and Gender. *The Stanford Encyclopedia of Philosophy*. E.N. Zalta, ed. URL <https://plato.stanford.edu/archives/fall2017/entries/feminism-gender/>
- Mills, Charles. 2000. But What Are You Really? *Radical Philosophy Today*. 1: 23-51.
- Montero, Barbara. 1999. The Body Problem. *Noûs*. 33(2): 183-200.
- Moore, G.E. 1903. *Principia Ethica*. Dover.
- Morris, Kevin. 2018. *Physicalism Deconstructed: Levels of Reality and the Mind-Body Problem*. Cambridge: Cambridge University Press.
- Nagel, Ernest. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace, and World.
- Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.
- Nei, M. and A.K. Roychoudhury. 1972. Gene Differences Between Caucasian, Negro, and Japanese Populations. *Science*. 177: 434-435.
- Ney, Alyssa. 2010. Are There Fundamental Intrinsic Properties? In *New Waves in Metaphysics*. A. Hazlett, ed. London: Palgrave-Macmillan: 219-239.
- Ney, Alyssa. 2021a. *The World in the Wave Function: A Metaphysics for Quantum Physics*. New York: Oxford University Press.
- Ney, Alyssa. 2021b. The Fundamentality of Physics: Completeness or Maximality. *Oxford Studies in Metaphysics*. 12.
- Nietzsche, Friedrich. *The Genealogy of Morals*.
- Nietzsche, Friedrich. *The Gay Science*.
- Office of Management and Budget (OMB). 2016. Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity. [www.regulations.gov/document?D=OMB-2016-0002-0001](http://www.regulations.gov/document?D=OMB-2016-0002-0001)

- Oppenheim, Paul and Hilary Putnam. 1958. The Unity of Science as a Working Hypothesis. *Minnesota Studies in the Philosophy of Science*. 2: 3-36.
- Papineau, David. 2012. *Philosophical Devices*. Oxford: Oxford University Press.
- Paul, L.A. 2006. In Defense of Essentialism. *Philosophical Perspectives*. 20: 333-372.
- Paul, L.A. 2012. Metaphysics as Modeling: The Handmaiden's Tale. *Philosophical Studies*. 160(1): 1-29.
- Pearl, Judea. 2000. *Causality*. Cambridge: Cambridge University Press.
- Peirce, C.S. 1877. The Fixation of Belief. *Popular Science Monthly*. 12(1): 1-15.
- Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.
- Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.
- Pereboom, Derk. 2021. *Wrongdoing and the Moral Emotions*. Oxford: Oxford University Press.
- Pereboom, Derk. 2022. *Free Will*. Cambridge: Cambridge University Press.
- Pink, Thomas. 2004. *Free Will: A Very Short Introduction*. Oxford: Oxford University Press.
- Place, U.T. 1956. Is Consciousness a Brain Process? *British Journal of Psychology*. 47: 44-50.
- Plantinga, Alvin. 1992. *The Nature of Necessity*. Oxford: Clarendon.
- Pojman, Louis P. and Michael C. Rea. 2007. *Philosophy of Religion: An Anthology* (5th Edition). Belmont: Wadsworth.
- Polger, Thomas and Lawrence Shapiro. 2016. *The Multiple Realization Book*. New York: Oxford University Press.
- Putnam, Hilary. 1967a. Psychological Predicates. In *Art, Mind, and Religion*. W.H. Capitan and D.D. Merrill, eds. Pittsburgh: University of Pittsburgh Press, 37-48.
- Putnam, Hilary. 1967b. Time and Physical Geometry. *Journal of Philosophy*. 64(8): 240-247.
- Putnam, Hilary. 1975. The Meaning of "Meaning." *Minnesota Studies in the Philosophy of Science*. 7: 131-193.
- Putnam, Hilary. 1979. "Philosophy of Logic." *Mathematics, Matter, and Method*. Cambridge: Cambridge University Press, 323-357.
- Quine, W.V. 1948. On What There Is. *Review of Metaphysics*. 2(5): 21-38. (Page references in text are to the version published in Quine. 1980. *From a Logical Point of View* (Second Revised Edition).) Cambridge: Harvard University Press.
- Quine, W.V. 1950. Identity, Ostension, and Hypostasis. *Journal of Philosophy*. 47: 621-633.
- Quine, W.V. 1951a. Two Dogmas of Empiricism. *Philosophical Review*. 60(1): 20-43.
- Quine, W.V. 1951b. On Carnap's Views on Ontology. *Philosophical Studies*. 2(5): 65-72.
- Quine, W.V. 1960. *Word and Object*. Cambridge: MIT Press.
- Quine, W.V. 1969. Natural Kinds. *Ontological Relativity and Other Essays*. New York: Columbia University Press, 114-138.
- Quine, W.V. 1981. *Theories and Things*. Cambridge: Harvard University Press.
- Raven, Michael, ed. 2021. *The Routledge Handbook of Metaphysical Grounding*. London: Routledge.
- Rea, Michael, ed. 1997. *Material Constitution: A Reader*. Oxford: Rowman and Littlefield.
- Reichenbach, Hans. 1956. *The Direction of Time*. Berkeley: University of California Press.
- Rich, Adrienne. 1977. *Of Woman Born*. New York: Bantam.
- Rosen, Gideon. 1990. Modal Fictionalism. *Mind*. 99(395): 327-354.
- Rosen, Gideon. 1994. Objectivity and Modern Idealism: What is the Question? In J. O'Leary-Hawthorne and M. Michael, eds. *Philosophy in Mind*. Dordrecht: Kluwer, 277-319.
- Rosen, Gideon. 2010. Metaphysical Dependence: Grounding and Reduction. In B. Hale and A. Hoffmann, eds. *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press.

- Rosenberg, Noah et al. 2002. Genetic Structure of Human Populations. *Science* 298(5602): 2381-2385.
- Routley, Richard. 1982. On What There Is Not. *Philosophy and Phenomenological Research*. 43(2): 151-177.
- Russell, Bertrand. 1912. *The Problems of Philosophy*. Oxford: Oxford University Press.
- Russell, Bertrand. 1927. *Analysis of Matter*. London: Kegan Paul.
- Ryle, Gilbert. 1949. *The Concept of Mind*. London: Hutchison.
- Salmon, Wesley. 1994. Causality Without Counterfactuals. *Philosophy of Science*. 61(2): 297-312.
- Sartorio, Carolina. 2005. Causes as Difference-Makers. *Philosophical Studies*. 123(1-2): 71-96.
- Sartorio, Carolina. 2016. *Causation and Free Will*. Oxford: Oxford University Press.
- Sartre, Jean-Paul. 1946/2007. *Existentialism is a Humanism*. In *Existentialism is a Humanism*. New Haven: Yale University Press.
- Schaffer, Jonathan. 2003. Is There a Fundamental Level? *Noûs*. 37(3): 498-517.
- Schaffer, Jonathan. 2004. Two Conceptions of Sparse Properties. *Pacific Philosophical Quarterly*. 85(1): 92-102.
- Schaffer, Jonathan. 2009. On What Grounds What. In *Metametaphysics*. D. Chalmers, D. Manley, and R. Wasserman, eds. Oxford: Oxford University Press.
- Schaffer, Jonathan. 2010. Monism: The Priority of the Whole. *Philosophical Review*. 119 (1): 31-76.
- Schaffer, Jonathan. 2015. What Not to Multiple without Necessity. *Australasian Journal of Philosophy*. 93: 644-664.
- Schaffer, Jonathan. 2016. Grounding in the Image of Causation. *Philosophical Studies*. 173(1): 49-100.
- Schaffer, Jonathan. 2017. The Ground Between the Gaps. *Philosophers' Imprint*. 17.
- Sellars, Wilfrid. 1962. Philosophy and the Scientific Image of Man. In *Frontiers of Science and Philosophy*. R. Colodny, ed. Pittsburgh: University of Pittsburgh Press, 35-78.
- Shoemaker, Sydney. 1998. Causal and Metaphysical Necessity. *Pacific Philosophical Quarterly*. 79 (1): 59-77.
- Shoemaker, Sydney. 2007. *Physical Realization*. Oxford: Oxford University Press.
- Sider, Theodore. 1993. Van Inwagen and the Possibility of Gunk. *Analysis*. 53(4): 285-289.
- Sider, Theodore. 1996. All the World's a Stage. *Australasian Journal of Philosophy*. 74(3): 433-453.
- Sider, Theodore. 2001. *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press.
- Sider, Theodore. 2003. Maximality and Microphysical Supervenience. *Philosophy and Phenomenological Research*. 66(1): 139-149.
- Sider, Theodore. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.
- Sider, Theodore. 2013. Against Parthood. *Oxford Studies in Metaphysics*. 8: 237-293.
- Sider, Theodore, John Hawthorne, and Dean Zimmerman. 2008. *Contemporary Debates in Metaphysics*. Oxford: Blackwell.
- Siegel, Susanna. 2010. *The Contents of Visual Experience*. Oxford: Oxford University Press.
- Sklar, Lawrence. 1974. *Space, Time, and Spacetime*. Berkeley: University of California Press.
- Skow, Bradford. 2015. *Objective Becoming*. Oxford: Oxford University Press.
- Slote, Michael. 1982. Selective Necessity and the Free Will Problem. *Journal of Philosophy*. 79(1): 5-24.
- Smart, J.J.C. 1959. Sensations and Brain Processes. *Philosophical Review*. 68: 141-156.
- Smart, J.J.C. 1963. *Philosophy and Scientific Realism*. London: Routledge.
- Smith, Barry and John Searle. 2003. An Illuminating Exchange: *The Construction of Social Reality*. *The American Journal of Economics and Sociology*. 62(1): 285-309.

- Sosa, Ernest and Michael Tooley. 1993. *Causality*. Oxford: Oxford University Press.
- Spelman, Elizabeth. 1988. *Inessential Woman: Problems of Exclusion in Feminist Thought*. Boston: Beacon Press.
- Spencer, Quayshawn. 2018. A Racial Classification for Medical Genetics. *Philosophical Studies*. 175(5): 1013-1037.
- Stalnaker, Robert. 1968. A Theory of Conditionals. In *Studies of Logical Theory*. N. Rescher, ed. Oxford: Blackwell.
- Stalnaker, Robert. 1976. Possible Worlds. *Noûs*. 10(1): 65-75.
- Stoljar, Natalie. 1995. Essence, Identity, and the Concept of Woman. *Philosophical Topics*. 23(2): 261-293.
- Strawson, Galen. 1989. *The Secret Connexion: Causation, Realism, and David Hume*. Oxford: Oxford University Press.
- Strawson, P.F. 1959. *Individuals: An Essay in Descriptive Metaphysics*. London: Routledge.
- Strawson, P.F. 1962. Freedom and Resentment. *Proceedings of the British Academy*. 48: 187-211.
- Strevens, Michael. 2007. Mackie Remixed. In *Causation and Explanation*. J. Campbell, M. O'Rourke, and H. Silverstein, eds. Cambridge: MIT Press.
- Swoyer, Chris. 1982. The Nature of Natural Laws. *Australasian Journal of Philosophy*. 60 (3): 203-223.
- Szabó, Zoltan Gendler. 2003. Nominalism. In *The Oxford Handbook of Metaphysics*. M. Loux and D. Zimmerman, eds. Oxford: Oxford University Press.
- Tahko, Tuomas. 2018. Fundamentality. *The Stanford Encyclopedia of Philosophy*. E.N. Zalta, ed. URL <https://plato.stanford.edu/archives/fall2018/entries/fundamentality/>
- Tallant, Jonathan. 2009. Presentism and Truth-making. *Erkenntnis*. 71(3): 407-416.
- Thomasson, Amie. 2007. *Ordinary Objects*. Oxford: Oxford University Press.
- Thomasson, Amie. 2015. *Ontology Made Easy*. Oxford: Oxford University Press.
- Thomasson, Amie. 2017. What Can We Do, When We Do Metaphysics? In *The Cambridge Companion to Philosophical Methodology*. G. d'Oro and S. Overgaard, eds. Cambridge: Cambridge University Press.
- Timpe, Kevin. 2016. Leeway vs. Sourcehood Conceptions of Free Will. In *Routledge Companion to Free Will*. London: Routledge, 213-224.
- Tooley, Michael. 1990. Causation: Reductionism vs. Realism. *Philosophy and Phenomenological Research*. 50: 215-236.
- Trogon, Kelly. 2013. An Introduction to Grounding. In *Varieties of Dependence*. M. Hoeltje, B. Schnieder, and A. Steinberg, eds. Munich: Philosophia Verlag.
- Unger, Peter. 1979. I Do Not Exist. In *Perception and Identity*. Graham Macdonald, ed. Ithaca: Cornell University Press.
- Unger, Peter. 1980. The Problem of the Many. *Midwest Studies in Philosophy*. 5(1): 411-468.
- Van Fraassen, Bas. 1989. *Laws and Symmetry*. Oxford: Oxford University Press.
- Van Inwagen, Peter. 1986. *An Essay on Free Will*. Oxford: Oxford University Press.
- Van Inwagen, Peter. 1990. *Material Beings*. Ithaca: Cornell University Press.
- Van Inwagen, Peter. 1998. Meta-ontology. *Erkenntnis*. 48(2-3): 233-250.
- Van Inwagen, Peter. 2000a. Temporal Parts and Identity across Time. *The Monist*. 83(3): 437-459.
- Van Inwagen, Peter. 2000b. Free Will Remains a Mystery. *Philosophical Perspectives*. 14: 1-19.
- Van Inwagen, Peter. 2009. Being, Existence, and Ontological Commitment. In *Meta-metaphysics*. D. Chalmers, Manley, and Wasserman (2009).
- Van Inwagen, Peter. 2017. Ability. In *Thinking About Free Will*. Cambridge: Cambridge University Press.
- Wang, Jennifer. 2016. Fundamentality and Modal Freedom. *Philosophical Perspectives*. 30: 397-418.
- Wasserman, Ryan. 2017. *Paradoxes of Time Travel*. Oxford: Oxford University Press.



- Wegner, Daniel. 2003. *The Illusion of Conscious Will*. Cambridge: MIT Press.
- Wiggins, David. 2001. *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.
- Wilkerson, Isabel. 2020. *Caste: The Origins of Our Discontents*. New York: Random House.
- Williams, D.C. 1951. The Myth of Passage. *Journal of Philosophy*. 48(15): 457-472.
- Williams, D.C. 1953. On the Elements of Being: I. *Review of Metaphysics*. 7(1): 3-18.
- Williamson, Timothy. 1994. *Vagueness*. London: Routledge.
- Wilson, Alistair. 2020. *The Nature of Contingency: Quantum Physics as Modal Realism*. Oxford: Oxford University Press.
- Wilson, Jessica. 1999. How Superduper Does a Physicalist Supervenience Need to Be? *Philosophical Quarterly*. 49(194): 33-52.
- Wilson, Jessica. 2005. Supervenience-based Formulations of Physicalism. *Noûs*. 39(3): 426-459.
- Wilson, Jessica. 2006. On Characterizing the Physical. *Philosophical Studies*. 131(1): 61-99.
- Wilson, Jessica. 2010. What is Hume's Dictum, and why believe it? *Philosophy and Phenomenological Research*. 80(3): 595-637.
- Wilson, Jessica. 2013. A Determinable-Based Account of Metaphysical Indeterminacy. *Inquiry*. 56(4): 359-385.
- Wilson, Jessica. 2014. No Work for a Theory of Grounding. *Inquiry*. 57(5-6): 535-579.
- Witt, Charlotte. 2011. *The Metaphysics of Gender*. Oxford: Oxford University Press.
- Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*. London: Kegan Paul.
- Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. Oxford: Blackwell.
- Wolf, Susan. 1990. *Freedom within Reason*. Oxford: Oxford University Press.
- Woodward, James. 2003. *Making Things Happen*. Oxford: Oxford University Press.
- Zack, Naomi. 2002. *Philosophy of Science and Race*. London: Routledge.
- Zack, Naomi. 2017. How Mixed Race is Not Constructed: US Identities and Perspectives. *The Oxford Handbook of Philosophy and Race*. Oxford: Oxford University Press.
- Zimmerman, Dean. 2008. The Privileged Present: Defending an 'A-theory' of Time. In *Contemporary Debates in Metaphysics*. T. Sider, D. Zimmerman, and J. Hawthorne, eds. Oxford: Blackwell, 211-225.
- Zimmerman, Dean. 2011. Presentism and the Space-time Manifold. *The Oxford Handbook of Philosophy of Time*. Oxford: Oxford University Press, 163-246.

# Index

Note: page numbers in italic type refer to Figures; those in bold type refer to Tables.

- 10 Downing Street example 101, 103
- a posteriori method of knowing 60, 61
- a priori method of knowing 60, 61
- abortion argument 11-13
- absence, of physical processes 323-324
- absolute fundamentality 201-202, 203, 210-211, 213-214, 215-216
- abstract, definition of 55
- abstract entities 41, 54-56; abstract/  
concrete distinction 56-59, **57**;  
further reading 81-82; mathematical  
objects 76-81, 80; nominalism 71-75;  
Quine's method 65-71; universals/  
One Over Many argument 59-65
- abstraction, definition of 58
- abundant ontology 47
- abundant theory of universals 64
- actualism 27, 289
- agent causal libertarianism 347-348
- agent causation 307
- Akiba, Ken 111-112
- Alcoff, Linda Martín 174-175, 176-177
- alien life forms argument 4-5
- alternative possibilities (or leeway)  
concept of freedom 340
- ameliorative approach 156, 159, 178,  
179
- analytic means of verification 118, 119,  
127
- "Analytical Language of John Wilkins,  
The" (Borges) 137, 137
- Andreason, Robin 160, 161, 162
- Anscombe, Elizabeth 147
- anti-essentialism 299-300
- anti-racist perspective 159
- Appiah, Anthony 160, 161
- "Are We Free to Break the Laws?"  
(Lewis) 337
- argument 1-3; criticizing 9-11; first-  
order predicate logic 18-27, **25**, 36,  
68-69, 76, 282; principle of charity/  
enthymemes 11-13; propositional  
logic 14-17, **15**; soundness 8-9;  
validity 4-6, **7**, 9, 10
- argument forms: predicate logic 18, 27;  
propositional logic 14-17
- Argument from Special Relativity 224,  
229-232
- Aristotle 299, 306; *Categories*  
186-187; *De Interpretatione* 60,  
61; *Metaphysics* 299; ontological  
commitment 38; *Physics* 257-258
- arithmetic: frameworks of 122, 123  
see also mathematics
- Armstrong, David 61, 64, 66, 67, 68, 70,  
73, 74, 213, 339
- Àsta 147, 151-153, 154, 180-181
- asymmetry 196
- atheism 2
- A-theory of time 220, 221, 222, 223-237,  
240-241, 251, 253
- 'atomic' propositions 14
- atomic theory of matter 77-78



- attorney, category of 146, 149-150, 151  
 Ayer, A.J. 126, 339-340, 341, 344
- Barnes, Elizabeth 111  
 basic propositions 14  
 Beauvoir, Simone de 170, 171, 174  
 beliefs: commonly-held 50-51; content of, as sets of possible worlds 287  
 Benacerraf, Paul 79-81, 80  
 Bennett, Karen 199, 200, 202, 204, 206, 207  
 Bettcher, Talia 172, 176  
 Big Bang argument 5-6  
 biological classifications 137, 138, 139, 159-160, 163  
 biological determinism, and gender 171-172  
 biological realism: about gender 158, 170, 171-174; about race 158, 159-163, 164, 165  
 biological species example 41-43  
 Bird, Alexander 140-141  
 Black Lives Matter movement 165  
 Black, Max 91  
 block universe view 221, 223, 224, 227, 230, 237, 249  
 Boolos, George 70  
 Borges, Jorge Luis 137, 137  
 bound variables 22  
 Boyd, Richard 141-143, 150  
 Brutal Composition 103, 109  
 B-theory of time 220-221, 223-237, 240, 241, 249, 252, 253  
 building relations 199, 204, 206-207, 212, 234  
 Butler, Judith 170
- Carnap, Rudolf 117-125, 199, 281  
 categorical features 238-239  
*Categories* (Aristotle) 186-187  
 category mistakes 186  
 causation: counterfactual theory 313, 313-317, 314, 316, 324, 325, 326; empiricism 308-311; further reading 326-327; historical perspectives 305-307, **308**; objects and events 307, **308**; primitivist theory 320-322, 325; probabilistic theory 317-320, 325; process theories 322-324, 325; regularity theory 310, 311-313, 314, 317, 325; relata of causal relations 306-307  
 "Causation" (Lewis) 315, 317  
 chains of verification 119, 119, 128  
 Chalmers, David 196  
 change, puzzle of 256-259, 270  
 charity, principle of 11-13  
 chemical elements, classification of 139-140  
 Chisholm, Roderick 347-348, 351  
 Church, Alonzo 44  
 class (set) nominalism 71-74, 75, 285-286  
 classifications 144; biological 137, 138, 139, 159-160, 163; chemical elements 139-140; racial 164, 165, 166, 168  
 cluster concept of gender 175-176  
 coextension, objection from 72, 286  
 cohesion 102, 109  
 coinciding objects 264  
 "color blind" approach 164  
 common sense 50-51  
 compatibilism 333, 337-342, 344-346, 353  
 "Concepts of Supervenience" (Kim) 191-192, 194-195  
 conceptualism 62  
 conclusions 2-3; major 10; minor 10, 12; soundness 8-9; validity 4-6, **7**, 9, 10  
 concrete entities 56-59, **57**  
 concrete material objects 83, 85-86  
 conferralism 151; gender 180-181  
 Consequence Argument 331-333, 339  
 constraints 151-152, 153; and free will 330, 331, 340, 343  
 contact 101-102, 109  
 contingency 284  
 contingent a priori 297, 298  
 contradiction 5  
 conventionalism 296-298  
 "cool kids" category 151  
 counterexamples 5  
 counterfactual theory of causation 313, 313-317, 314, 316, 324, 325, 326

- counterfactuals 284-285
- counterparts 271
- criticizing arguments 9-11
- critiques of metaphysics: Carnap
  - 117-125; further reading 134-135;
  - methodological concerns 116-117,
  - 130-132; Quine 127-130; relationship
  - between metaphysics and science
  - 132-134; responses to Carnap
  - 125-130
- Cube Factory 318-319
- cultural constructionism, and race
  - 168-169
- Curry, Stephen 34, 38
  
- Dalton, John 107
- Daly, Chris 198
- Davidson, Donald 44-46, 307
- de dicto modality 280, 298, 300, 302
- de re modality 280, 298, 299, 300,
- 302
- deductive argument 214
- deductive validity/invalidity 4-6, **7**;
  - criticizing arguments 9-11; principle
  - of charity/enthymemes 11-13
- definite descriptions, names as 37-38
- Descartes, René 84, 185, 299, 348
- descriptions, definite and indefinite
  - 37-38
- descriptive metaphysics 154
- design, argument from 2, 9-11
- determinism 246-247, 331-337, 334,
- 340, 347; hard determinism 334,
- 353
- diachronic identity 96
- dilemma 79
- dinosaurs example 235, 239-241, 243,
- 244, 245, 246, 247-248
- disability 152
- disjunctive syllogism 15, 16
- dispositional features 238
- domains of quantification 20, 22-23,
- 25, 27
- donuts example 40-41
- Dowe, Phil 322-323, 324, 325-326
- DuBois, W.E.B. 168
  
- efficient cause 306
- Einstein, Albert 224, 224-226, 225, 227,
- 231
- Eleatic principle 165, 176
- electrons examples 26, 39-40, 43
  - see *also* particle physics
- eliminativism 146; about gender 158,
- 159, 171, 176-177, 181-182; about race
- 158, 159, 164-166
- empiricism 308-311
- "Empiricism, Semantics, and Ontology"
- (Carnap) 121-124
- endurantism 259, 259, 260-261, 262,
- 265-266, 267, 272
- Enquiry Concerning Human*
- Understanding* (Hume) 309
- enthymemes 11-13
- entrenchment 150
- epiphenomena 313, 313, 319-320
- epistemicism 110
- epistemic 155
- epistemic possibility 331
- epistemological fundamentality 201
- epistemology 79; of mathematics
- 80-81
- equivalence relations 88
- ersatz modal realism 288, 289-294
  - see *also* linguistic ersatzism
  - (linguistic ersatz modal realism)
- essence 139, 141
- essential properties 280, 288-289
- essentialism: about gender 174-176;
- and modality 298-303; and race
- 169
- eternal time 249
- eternalism 221, 222, 223, 224, 232-233,
- 249, 253, 260, 266; truthmaker
- theory 237, 240, 241, 248
- event causal libertarianism 348-349,
- 351
- events, existence of 44-46
- evil 3-4, 145
- exdurantism (stage theory) 267-268,
- 269-270, 269-271, 275
- existential quantifiers 18-21, 24-26, **25**,
- 33, 69, 71

- existentialism 349-350
- extended simples 274
- external questions 122
- external statements 122, 123
  
- fastening 102, 109
- Fausto-Sterling, Anne 173
- Feigl, Herbert 188
- feminist theory 159, 174-175
- fictionalism 244-245, 295-296
- Field, Hartry 78-79, 124
- final (teleological) causes 306
- Fine, Kit 51, 302
- first-order predicate logic 18-27, **25**, 36, 68-69, 76, 282
- first-person authority over gender status 172-173
- Fischer, John Martin 341
- Fleabert example 101, 103, 108
- Floyd, George 165
- Fodor, Jerry 189
- form 60, 61, 258, 306
- Forrest, Peter 289
- foundationalism 201
- four dimensionalism 261, 262, 267, 272, 274-275
- frameworks 121-122
- Frankfurt, Harry 341, 342-343
- free will 14-15, 329-331; compatibilism 333, 337-342, 344-346, 353; determinism 331-337, 334, 340, 347; Frankfurt cases 342-343; further reading 354; libertarianism 334, 347-351; manipulation arguments 343-347; and moral responsibility 352-353; skepticism about 351-353
- Frege, Gottlob 18
- Friedan, Betty 174
- fundamentality 184-185, 200-202; as basis for nonfundamentality 210-213; fundamental level 213-216; further reading 216-217; grounding 196-200, 204, 206-207, 206-210, 213, 234; as a mereological notion 205, 205-206; mind-body problem 185-190; as ontological independence 202-210, 212; realization 188-190; supervenience 190-197, 198 *see also* mind-body problem
- fusion 102, 109; mereological sum 99
  
- gender 154, 169-171; biological realism 158, 170, 171-174; cluster concept of 175-176; conferralism 180-181; eliminativism 158, 159, 171, 176-177, 181-182; essentialism 174-176; first-person authority over gender status 172-173; further reading 182; gender identities 178-179, 181; positional view 176-177; social constructionism 158-159, 171, 177-180
- genotypes 160
- Glasgow, Joshua 162, 167
- global supervenience 192
- God 2-3, 195-196, 204, 207
- Goliath *see* Statue and the Clay paradox
- good, concept of 145
- Greece in EU argument 8-9
- grounding 196-200, 204, 206-207, 206-210, 213; further reading 216-217; and identity 200
- growing block theory 222
  
- Hacking, Ian 145-146, 147-148, 149, 150, 151, 152, 154
- hard determinism 334, 353
- hard incompatibilism 351, 353
- Harris, Sam 351, 352
- Harry Potter example 242-243, 245
- Haslanger, Sally 149, 151, 156, 166-167, 177-178, 179-180, 181, 266-267, 274
- Hawley, Katherine 140-141
- healthcare argument 3
- Heidegger, Martin 120
- Heil, John 213
- Hobbes, Thomas 340
- Hofweber, Thomas 198
- homeostatic mechanism 142
- hooks, bell 159
- Horwich, Paul 249, 251, 253

- human agents 307
- human bodies example 96-97, 112, 258-259
- human kinds, construction of 147-153
- Hume, David 305, 308-310, 339
- Humeanism about laws 338-339
- hylomorphism 258
  
- idealism 27, 185
- idealizations 78
- identity 27; gender 178-179, 181; and grounding 200; of indiscernables 91; *No entity without identity* (Quine) 33-34; numerical identity (identity in the strict sense) 34; qualitative 35
- identity theory of mind 188
- image, manifest and scientific 50, 51
- immanent entities 60, 61
- incompatibilism 333, 338, 342, 343, 346, 351
- indefinite descriptions 37-38
- indeterminism 335-337, 348
- indiscernability of identicals (Leibniz's Law) 90-92, 93, 113, 256, 257, 266, 267, 272
- indispensability argument 72, 76-79, 81, 285
- inductive argument 214
- inertial reference frames 224, 225
- inference: principle of charity/enthymemes 11-13; validity 10
- in-groups 149
- instantiation 59
- interactive kinds 147-148
- interactive substance dualism 185
- internal maps, gender identities as 179
- internal questions 122
- internal statements 122, 123
- intersectionality 177
- intersubjectivity 144
- intrinsic properties 265, 267
- intuitions 50-51, 108, 113, 131
- invalid arguments see validity
- irreflexivity 196
  
- Jackson, Frank 70-71
- Jeffers, Chike 168-169, 181
- Jenkins, Katherine 178-179, 179-180
  
- Kane, Robert 330, 348-351
- Kenny, Anthony 45
- Kim, Jaegwon 190, 191-192, 194-195, 198, 307
- Koslicki, Kathrin 198-199
- Kripke, Saul 297-298, 300, 301
  
- Ladyman, James 50, 131-132
- language, distortions of (Carnap) 117-121
- Laplace, Pierre-Simon 334-335
- law of excluded middle 129
- leeway (or alternative possibilities) concept of freedom 340
- Leibniz, Gottfried 90, 195, 282
- Leibniz's Law (indiscernability of identicals) 90-92, 93, 113, 256, 257, 266, 267, 272
- Leviathan* (Hobbes) 340
- Lewis, C.I. 281
- Lewis, David 73, 74, 107, 110, 141, 249, 252, 253, 265-266, 267-268, 268, 283-284, 285, 287-288, 289, 290, 294, 295, 325, 326; counterfactual theory of causation 313, 313-317, 314, 316; free will 337-338, 339; "New Work for a Theory of Universals" 339; *On the Plurality of Worlds* 107, 265-266, 267-268, 283-284, 285, 295
- libertarianism 334, 347-351
- Libet, Benjamin 352
- linguistic ersatzism (linguistic ersatz modal realism) 289-294
- linguistic vagueness 110
- local supervenience 193
- Locke, John 299, 342
- logic: first-order predicate logic 18-27, **25**, 36, 68-69, 76, 282; modal 281-282; propositional 14-17, **15**, 281; second-order predicate logic 68-70
- logical connectives 15
- logical positivism 117, 118, 122, 125, 126, 131, 299
- logical possibility 278-279, 279
- logicism 118
- Lowe, E.J. 187
- Lump see Statue and the Clay paradox

- Maddy, Penelope 77-78  
 major conclusions 10  
 Mallon, Ron 146-147, 148-150, 151, 152, 153-154  
 manifest image 50, 51  
 manipulation arguments 343-347  
 Marcus, Ruth Barcan 281, 302  
 Markosian, Ned 103, 108  
 Martin, Trayvon 165  
 mass, Newton's definition 85  
 material objects 55, 84-86;  
   mereological universalism 107-109;  
   paradoxes of material constitution 86, 86-96, 93, 112-113, 256-257, 261-264, 272; Problem of the Many 86, 96-98, 112; Special Composition Question 86, 99-113; vagueness 109, 110-112  
 materialism 185  
 mathematics 124-125; epistemology of 80-81; mathematical objects 55, 76-81, 80; set theory 72  
 matter 77-78, 258, 306  
 Maudlin, Tim 339  
 maximal proximity 97-98  
 McDaniel, Kris 274  
 McKenna, Michael 345-346  
 McTaggart, J.M.E. 52n5, 233, 236, 237  
 McX (Quine's character) 31-32, 35, 62, 63, 67, 155, 242  
 meaning: distortions of (Carnap) 117-121; verificationist theory of 117-119, 119, 121, 122, 125-126, 130-131  
 Meinong, Alexius 52n5  
 Mele, Alfred 346  
 Mellor, D.H. 307  
 Melynk, Andrew 200  
 mental properties 186-191, 192, 194, 195-196, 198, 200-201, 202-203  
 mereological atom (simple) 100, 104, 105, 106, 107  
 mereological nihilism 105-107, 112, 124  
 mereological sum (fusion) 99  
 mereological universalism 107-109, 112, 113, 262  
 meta-ontology 46, 76  
 metaphysical explanation 68  
 metaphysical fundamentality 201  
 metaphysical (ontic) vagueness 110  
 metaphysics: relationship with science 132-134; starting point for 48-51  
*Metaphysics* (Aristotle) 299  
*Metaphysics Within Physics, The* (Maudlin) 339  
 Mills, Charles 165, 166  
 mind-body dualism 106, 348  
 mind-body problem 50, 185-190, 193, 198, 200  
 Minkowski, Hermann 228, 228  
 minor conclusions 10, 12  
 mismatch objection 162, 180; gender 177-178  
 missing premises 13  
 modal claims 277-278  
 modal independence 202-204  
 modal logic 281-282  
 modal properties 94  
 modal realism 283-288, 289-294  
 modality: ersatz modal realism 288, 289-294; essentialism and anti-essentialism 298-303; further reading 303; possibility and necessity 277-280; possible worlds analysis 280-288, 287, 294-298  
 models 132-133  
 modes of truth 277-278  
 modus ponens 16  
 modus tollens 15, 16  
 monadic property 266  
 morality 145, 146; moral responsibility and free will 352-353  
 mortality argument 14-15, **15**  
 moving spotlight view 221, 222-223  
 multiple realization argument 188-190  
 multiverse 288  
  
 names, as definite descriptions 37-38  
 natural kinds 136-137, 137, 138, 139-141, 143-144, 154; further reading 156-157; Homeostatic Property Cluster account 141-144, 150  
 "Natural Kinds" (Quine) 137  
 natural properties 74, 141  
 naturalism 49, 50, 51, 76, 77, 78, 299, 306, 311  
 naturalist metaphysicians 132

- naturalistic dualism 196
- necessary a posteriori 297, 298
- necessity 277-280
- neuroscience 205-206
- "New Work for a Theory of Universals" (Lewis) 339
- Newton, Isaac 85, 133
- Newtonian physics 79, 226, 227, 334-335
- Nietzsche, Friedrich 145
- nihilism *see* mereological nihilism
- nomc regularity theory 312-313, 317
- nominalism 27, 62, 71, 77, 124, 125, 198; class (set) 71-74, 75; and natural kinds 140-141; ostrich 71; predicate 71, 75
- nomological possibility of necessity 278, 279
- nonexistent objects 30-35
- nonreductive physicalism 188
- nothingness 120
- numbered premise form 2
- numbers 55, 58, 76, 124, 204 *see also* mathematics
- numerical identity (identity in the strict sense) 34
  
- objection from coextension 72, 286
- Ockham's Razor 46-48, 105, 188
- Olympic Games 2028 example 235, 239-240, 241, 243-244, 245, 246, 247-248
- On the Plurality of Worlds* (Lewis) 107, 265-266, 267-268, 283-284, 285, 295
- "On What There Is" (Quine) 30, 36, 37, 41, 47, 62, 63, 65, 67, 71
- One Over Many argument 62-65, 66, 67, 69, 70, 71, 76
- ontic (metaphysical) vagueness 110
- ontological categories 186-187
- ontological independence, fundamentality as 202-210, 212
- ontological questions 122-124, 129-130
- ontology 29-30; existence of events 44-46; further reading 51-52; meta-ontology 46, 76; nonexistent objects 30-35; ontological commitments 30, 35-46; paraphrase method 41-44; parsimony in (Ockham's Razor) 46-48; starting points for metaphysical enquiry 48-51
- openness of the future 220
- Oppenheim, Paul 205-206
- origins essentialism 300-301
- ostrich nominalism 71
- out-groups 149
  
- painting example 193
- pandas example 237-238, 239, 247
- paraphrase method 41-44; Special Composition Question 104, 105; universals 70-71
- Parmenides 257
- particle physics 85; ontological commitment 39-40
- particulars 60
- passage of time 219-224
- Paul, Laurie 132-134
- Pegasus example 30-32, 34, 35-38, 155, 156
- perdurantism 187, 260, 260, 261, 262, 263-264, 264-265, 267, 269-270, 272, 275
- Pereboom, Derk 344-345, 346, 351, 353
- persistence 257-259; endurantism 259, 259, 260-261, 262, 265-266, 267, 272; exdurantism (stage theory) 267-268, 269-270, 269-271, 275; further reading 275; material constitution paradoxes 261-264; perdurantism 260, 260, 261, 262, 263-264, 264-265, 267, 269-270, 272, 275; puzzle of change 256-259, 270; temporary intrinsics problem 264-269, 268, 272; three dimensionalism 261, 272-275
- personal time 249
- phenomenalism 187
- phenotypes 160
- Physical Causation* (Dowe) 322-323
- physical properties 186, 187-188, 189-190, 191-192, 193-196, 198, 200-201, 203, 206

- physical theories of causation 322-324
- physicalism 49-50, 106, 111, 187-189, 190, 191-192, 193, 194-196, 200, 201, 202-203, 348
- physics 50, 77-78, 107, 139, 214 *see also* Newtonian physics; particle physics; quantum mechanics
- Physics* (Aristotle) 257-258
- Place, U.T. 188
- Plantinga, Alvin 289
- Plato 31, 60, 63, 64, 76, 137, 165, 185, 199
- Platonism 62, 76, 80, 81, 204, 280
- Polger, Thomas 200
- poor worldmaking language 290, 292-294
- positional view of gender 176-177
- possibilism 289
- possibility 277-280, 279
- possible worlds analysis 280-288, 287; rejection of 294-298
- precisificational worlds 111
- predicate logic: first-order 18-27, **25**, 36, 68-69, 76, 282; second-order 68-70
- predicate nominalism 71, 75
- preemption 314, 314, 319, 320
- preestablished harmony 195
- premises 2-3; soundness 8-9; supplying missing 13; validity 4-6, **7**, 9, 10
- pre-philosophical intuitions 50-51
- presentism 27, 221-222, 224, 232-233, 265-266; Argument from Special Relativity against 229-232; truthmaker theory 241-244, 245-246, 246-247, 248
- primitive concepts 198
- primitivist theory of causation 320-322, 325
- principle of charity 11-13
- Principle of Naturalistic Closure 131-132
- Prior, Arthur 241
- priority monism 208, 210
- priority pluralism 206
- probabilistic theory of causation 317-320, 325
- process theories of causation 322-324, 325
- projectability 139
- proper parts 100
- properties 198, 285-286; abstract entities 55, 57-75; natural 74; sparse theory of 74
- property dualism 187
- propositional logic 14-17, **15**, 281
- propositions 56, 128
- protocol statements 119
- Putnam, Hilary 76, 189, 205-206, 297
- qualitative identity 35
- qualitative parsimony 48
- quantification, domains of 22-23, 25, 27
- quantifiers: existential 18-21, 24-26, **25**, 33, 69, 70; scope of 21-22; universal 23, 24, **25**, 26, 27
- quantitative parsimony 48
- quantum mechanics 129, 133, 335-337
- Queen Anne example 236, 237
- Queen of England (Elizabeth II) example 178, 180, 300-302
- Quine, W.V. 29-35, 49, 127-128, 155, 187, 199, 242, 260, 288, 299, 300, 301; abstract entities 62, 63, 65-71, 76, 79, 80; *To be is to be the value of a bound variable* 37, 40; natural kinds 137, 139; "Natural Kinds" 137; *No entity without identity* 33-34; "On What There Is" 30, 36, 37, 41, 47, 62, 63, 65, 67, 71; ontological commitments 35-46; ontological parsimony 47-48; paraphrase method 41-44; "Two Dogmas of Empiricism" 127-130; web of belief 128, 128-130
- race 154, 158-159; biological realism 158, 159-163, 164, 165; cultural constructionism 168-169; eliminativism 158, 159, 164-166; further reading 182; social constructionism 158-159, 164, 166-169
- racial classifications 164, 165, 166, 168; U.S. Census 160, 162, 163
- racialisation 166-167

- racialism 161
- racism 161, 164, 167, 168
- ravens argument 214-215
- realism 146, 198; about mathematical entities (Platonism) 76; about universals 62, 64, 66, 70, 71, 75, 125
- realization 189; multiple realization argument 188-190
- reasons-responsive compatibilism 241
- reductio ad absurdum 94-95, 172
- reductionism 127
- reductive physicalism 187-188
- reference class problem 318
- reflexivity 88, 196
- regimentation 36, 104-105
- regularity theory of causation 310, 311-313, 314, 317, 325
- Reichenbach, Hans 322
- relative fundamentality 185, 201-202, 210; grounding 196-200, 204, 206-207, 206-210, 213, 234; realization 188-190; supervenience 190-197, 198
- relativity objection 163
- Republic* (Plato) 63
- revisionary metaphysics 154
- Rich, Adrienne 174
- rich worldmaking language 290, 291-292, 294
- Rosen, Gideon 295
- Ross, Don 50, 131-132
- Russell, Bertrand 37, 38, 187
- Ryle, Gilbert 186
- Salmon, Wesley 322
- Santa Claus example 30-32, 35-38
- Sartorio, Carolina 325
- Sartre, Jean-Paul 349-350
- Schaffer, Jonathan 196, 198, 199-200, 202, 204, 208-209, 210, 213, 215
- Schrödinger, Erwin 133, 336
- science: relationship with metaphysics 132-134; scientific theory 49-50, 76-77
- scientific image 50, 51
- scope (of a quantifier) 21-22
- second-order predicate logic 68-70
- self-forming actions 349-351
- self-identifications 149
- Sellars, Wilfrid 50
- semantic ascent 36, 270
- semantic theory 79
- sense impressions 308-309, 310
- set (class) nominalism 71-74, 75, 285-286
- set theory 72
- sex 169-171 *see also* gender
- Shapiro, Lawrence 200
- Ship of Theseus paradox 86, 86-90, 92-93, 99, 112, 256, 257, 262, 262-263, 269-270
- Sider, Theodore 97-98, 214-215, 268, 269
- simple compatibilism 340-341, 344
- simple (mereological atom) 100, 104, 105, 106, 107
- simple regularity theory 311-312, 317 *see also* regularity theory of causation
- simplification 15, 16
- Smart, J.J.C. 188
- social construction 136, 144-147, 153; causal v. constitutive 153-154; constructing human kinds 147-153; further reading 156-157
- social constructionism: about gender 158-159, 171, 177-180; about race 158-159, 164, 166-169
- social kinds 136, 143-145, 154; further reading 156-157; genders as 171; races as 164, 166 *see also* social construction
- social ontology 154-156
- social processes 153-154
- social roles 149-150
- social significance 151
- social status 151
- soft determinism 334
- Solar System example 101-102
- sortal essentialism 301-302
- soundness, of arguments 8-9
- sourcehood conception of freedom 340
- space-time 228, 228-229, 229



- spacetime worms 269, 275
- sparse ontology 47, 73, 75
- sparse theory of properties 74
- sparse theory of universals 64, 73-74
- spatiality, of material objects 84-85
- Special Composition Question 99-100;
  - mereological nihilism 105-107, 112;
  - mereological universalism 107-109, 112, 113; moderate answers to 100-105, 109, 110, 113
- special relativity 224-229, 225, 226, 228, 229
- species, biological example 41-43
- Spencer, Quayshawn 160-161, 162
- stage theory (exdurantism) 267-268, 269-270, 269-271, 275
- Stage View 268, 269
- Stalnaker, Robert 285, 289
- Statue and the Clay paradox 93, 93-96, 99, 112, 113, 256-257, 259, 259-260, 260, 264
- Stoljar, Natalie 173, 175-176
- Strawson, P.F. 154, 155
- strong supervenience 192, 195, 196
- substance dualism 185
- Superman example 92
- supervenience 190-194, 198; limits of 194-197
- surface freedom 330
- symbolizations: first-order predicate logic 14; modal claims 282; propositional logic 14-16, **15**
- symmetry 88
- synchronic identity 96
- synthetic means of verification 118-119, 123, 127, 128
- Tahko, Tuomas 213-214
- teleological (final) causes 306
- temporal counterparts 271
- temporal fictionalism 244-245
- temporary intrinsics, problem of 264-269, 268, 272
- Thales of Miletus 13
- theism 2, 9-11, 204, 306
- theory of abstract particulars (trope theory) 74-75, 141, 187, 307
- thought experiments 133, 225-226, 226, 342
- three dimensionalism 261, 272-275
- time: A-theory 220, 221, 222, 223-237, 240-241, 251, 253; B-theory 220-221, 223-237, 240, 241, 249, 252, 253; further reading 253-254; material objects 84-85; passage of 219-224; special relativity 224-232, 225, 226, 228, 229; time travel 248-253, 250; truthmaker objection 237-248
- To be is to be the value of a bound variable* (Quine) 37, 40
- Tooley, Michael 320-322, 326
- Tractatus Logico-Philosophicus* (Wittgenstein) 126
- transcendent entities 56, 60, 61
- transgender people 172, 181; trans women 178, 180-181
- transitivity 88, 196
- trope theory (theory of abstract particulars) 74-75, 141, 187, 307
- truth, modes of 277-278
- truthmaker maximalism 237
- truthmaker theory 67-68, 76, 212-213, 234; and time 237-248; verificationist theory of 118
- "Two Dogmas of Empiricism" (Quine) 127-130
- Two Object View 95, 264
- two-dimensional time 250, 250-251
- two-placed predicate 19
- type materialism 187-188
- ultimate freedom 330
- Unger, Peter 86, 96, 97
- universal quantifiers 23, 24, **25**, 26, 27
- universalists (mereological universalists) 108-109
- universals 59-62; abundant theory of 64; applying Quine's method 65-71; and natural kinds 140-141; numbers as 76; One Over Many argument 62-65, 66, 67, 69, 70, 71, 76; paraphrasing away 70-71; sparse theory of 64

- U.S. Census, racial classification 160, 162, 163
- use/mention distinction 31
- vagueness 109, 110-112, 173
- validity: of arguments 4-6, **7**, 9, 10
  - see *also* deductive validity/invalidity
- van Fraassen, Bas 318-319
- van Inwagen, Peter 33, 50, 273, 274, 351; Consequence Argument 331-333, 338, 339; material objects 85-86; Special Composition Question 99-100, 102, 103-104, 105, 106-107, 112, 113
- variables 19; bound 22
- Venus example 270-271
- verification, methods of 118-119, 119
- verificationist theory of meaning 117-119, 119
- verificationist theory of truth 118
- Vienna Circle 117, 126
- Wang, Jennifer 202, 203
- water, as a natural kind 136, 137, 139, 140
- weak supervenience 192
- web of belief (Quine) 128, 128-130
- Wegner, Daniel 351
- What is a Law of Nature?* (Armstrong) 339
- wholly present 261
- widowhood example 149
- Wiggins, David 95, 113, 264, 272-273, 274
- Williams, D.C. 74, 75, 187, 248, 253
- Williams, J. Robert G. 111
- Wilson, Alistair 288
- Wilson, Jessica 85, 111, 198, 200, 206
- witches 164
- Wittgenstein, Ludwig 117, 126, 175, 277
- Wolf, Susan 341, 344
- "woman question" in feminist theory 174-175
- women: persecution as witches 164
  - see *also* gender
- worldline 323
- worldmaking language: poor 290, 292-294; rich 290, 291-292, 294
- Worm View 268, 269 see *also* perdurantism
- Wyman (Quine's character) 31-33, 35, 242
- Zimmerman, George 165
- Zygote Argument 346, 347